Statistics Canada Statistique Canada


# Census Recensement Canada 1986

## USER'S GUIDE TO THE QUALITY OF 1986 CENSUS DATA:

## SAMPLING AND WEIGHTING

Canada

**Canadian Cataloguing in Publication Data**

Main entry under title:

User's guide to the quality of 1986 census data : sampling and weighting

(Reference)
(Census Canada, 1986 = Recensement Canada, 1986)
Issued also in French under title: Guide à l'intention des utilisateurs sur la qualité des données du recensement de 1986, échantillonnage et pondération.
ISBN 0-660-13353-9
CS99-136E

1. Canada -- Census, 1986 -- Methodology. 2. Canada -- Census, 1986 -- Evaluation. I. Statistics Canada. II. Title: Guide à l'intention des utilisateurs sur la qualité des données du recensement de 1986, échantillonnage et pondération. III. Series: Reference. IV. Series: Census Canada, 1986 = Recensement Canada, 1986.

HA741.5.1986 U83 1990 304.6'0723 C90-098027-3

Statistics Canada Statistique Canada

Census
Recensement

Canada 1986

**Reference**

# USER'S GUIDE TO THE QUALITY OF 1986 CENSUS DATA:

# SAMPLING AND WEIGHTING

Published under the authority of the Minister of Industry, Science and Technology



October 1990

*Price:* Canada, $23.00
Other Countries, $24.00

Catalogue 99-136E

ISBN-0-660-13353-9

Ottawa

La version française est disponible sur demande (Catalogue Nº. 99-136F).

**Canadian Cataloguing in Publication Data**

Main entry under title:

User's guide to the quality of 1986 census data : sampling and weighting

(Reference)
(Census Canada, 1986 = Recensement Canada, 1986)
Issued also in French under title: Guide à l'intention des utilisateurs sur la qualité des données du recensement de 1986, échantillonnage et pondération.
ISBN 0-660-13353-9
CS99-136E

1. Canada -- Census, 1986 -- Methodology. 2. Canada -- Census, 1986 -- Evaluation. I. Statistics Canada. II. Title: Guide à l'intention des utilisateurs sur la qualité des données du recensement de 1986, échantillonnage et pondération. III. Series: Reference. IV. Series: Census Canada, 1986 = Recensement Canada, 1986.

HA741.5.1986 U83 1990 304.6'0723 C90-098027-3

**ACKNOWLEDGEMENTS**

This document was prepared by Stephen Rathwell of the Social Survey Methods Division.

The author would like to thank Mike Bankier, Pierre Daoust, Yves Béland, Don Royce, and Dick Carter of the Census Methodology and Research Subdivision for their input to this document.

The author is also most grateful to Louise Whalen and Lorna Bailie of the Census Operations Division for their contributions.

Parts of the first three chapters were taken from Quality of Data, Series 1: Sources of Error - Sampling and Weighting, the 1976 Census version of this document prepared by Gordon Brackstone.

## TABLE OF CONTENTS



**List of Tables**

**TABLE OF CONTENTS** - Concluded



**List of Figures**

## I. INTRODUCTION

Sampling is an accepted practice in many aspects of life today. The quality of produce in a market may be judged visually by a sample before a purchase is made; we form opinions about people based on samples of their behaviour; we form impressions about countries or cities based on brief visits to them. These are all examples of sampling in the sense of drawing inferences about the "whole" from information for a "part".

In a more scientific sense, sampling is used, for example, by accountants in auditing financial statements, in industry for controlling the quality of items coming off a production line, and by the takers of opinion polls and surveys in producing information about a population's views or characteristics. In general, the motivation to use sampling stems from a desire either to reduce costs or to obtain results faster, or both. In some cases, measurement may destroy the product (e.g., testing the life of light bulbs) and sampling is therefore essential. The disadvantage of sampling is that the results based on a sample may not be as precise as those based on the whole population. However, when the loss in precision (which may be quite small when the sample is large) is tolerable in terms of the uses to which the results are to be put, the use of sampling may be cost-effective. Furthermore, the reduction in the scale of a study achieved through using sampling may in fact lead to a reduction in errors from non-sampling sources, thus compensating to some extent for the loss of precision resulting from sampling.

The 1986 Census of Population made use of sampling in a variety of ways. It was used in ensuring that the quality of the Census Representative's work in collecting questionnaires met certain standards; it was used in the control of the quality of coding responses during office processing; it was used in estimating both the amount of under-coverage and the amount of over-coverage which occurred for different reasons; it was used in evaluating the quality of census data. However, the primary use of sampling in the census was during the field enumeration when all but the basic census data were collected only from a sample of households. This guide describes this last use of sampling and evaluates the effect of sampling on the quality of census data.

Chapter II reviews the history of the use of sampling in Canadian censuses and describes the sampling procedures used in the 1986 Census. Chapter III explains the procedures used for weighting up the sample data to the population level and provides operational and theoretical justifications for these procedures. In Chapter IV the program of studies designed to evaluate the 1986 Census sampling and weighting procedures is presented, while Chapters V through VIII present the results of these studies.

## II. SAMPLING IN CANADIAN CENSUSES

In the context of a Census of Population, sampling refers to the process whereby certain characteristics are collected and processed only for a random sample of the dwellings and persons identified in the complete census enumeration. Tabulations that depend on characteristics collected only on a sample basis are then obtained for the whole population by scaling up the results for the sample to the full population level. Characteristics collected on all dwellings or persons in the census will be referred to as "basic characteristics" while those collected only on a sample basis will be known as "sample characteristics".

### A. The History of Sampling in the Canadian Census[1]

Sampling was first used in the Canadian census in 1941. A Housing Schedule was completed for every tenth dwelling in each census subdistrict. The information from 27 questions on the separate Housing Schedule was integrated with the data in the personal and household section of the Population Schedule for the same dwelling, thus allowing cross-tabulation of sample and basic characteristics. Also in the 1941 Census, sampling was used at the processing stage to obtain early estimates of earnings of wage-earners, of the distribution of the population of working age, and of the composition of families in Canada. In this case, a sample of every tenth enumeration area across Canada was selected and all Population Schedules in these areas were processed in advance.

Again in 1951, the Census of Housing was conducted on a sample basis. This time every fifth dwelling (those whose identification numbers ended in a 2 or 7) was selected to complete a housing document containing 24 questions. In the 1961 Census, persons 15 years of age and over in a 20% sample of private households were required to complete a Population Sample Questionnaire containing questions on internal migration, fertility and income. Sampling was not used in the smaller censuses of 1956 and 1966.

The 1971 Census saw several major innovations in the method of census-taking. The primary change was from the traditional canvasser method of enumeration to the use of self-enumeration for the majority of the population. This change was prompted by the results of several studies in Canada and elsewhere (Fellegi (1964); Hansen et al. (1959)) that indicated that the effect of the enumerator was a major contribution to the variance[2] of census figures in a canvasser census. Thus the use of self-enumeration was expected to reduce the variance of census figures through reducing the effect of the enumerator, while at the same time giving the respondent more time and privacy in which to answer the census questions - factors which might also be expected to yield more accurate responses.

---

[1] More detailed information for specific censuses can be found in the Administrative Report, General Review, Summary Guide or Census Handbook of the appropriate census. References to these reports can be found at the end of this guide.

[2] The "variance" of an estimate is a measure of its precision. Variance is discussed more fully in Chapter VIII.

The second aspect of the 1971 Census that differentiated it from any earlier census was its content. The number of topics covered and the number of questions asked were greater than in any previous Canadian census. Considerations of cost, respondent burden, and timeliness versus the level of data quality to be expected using self-enumeration and sampling led to a decision to collect all but certain basic characteristics on a one-third sample basis in the 1971 Census. In all but the more remote areas of Canada, every third private household received the "long form" which contained all the census questions, while the remaining private households received the "short form" containing only the basic questions covering name, relationship to head, sex, date of birth, marital status, mother tongue, type of dwelling, tenure, number of rooms, water supply, toilet facilities, and certain coverage items. All households in pre-identified remote enumeration areas and all collective dwellings[3] received the long form. A more detailed description of the consideration of the use of sampling in the 1971 Census is given in Sampling in the Census (Dominion Bureau of Statistics (1968)).

The content of the 1976 Census was considerably less than that of the 1971 Census. Furthermore, the 1976 Census did not include the questions that cause the most difficulty in collection (e.g., income) or that are costly to code (e.g., occupation, industry, and place of work). Therefore, the benefits of sampling in terms of cost savings and reduced respondent burden were less clear than for the 1971 Census. Nevertheless, after estimating the potential cost savings to be expected with various sampling fractions, and considering the public relations issues related to a reversion to 100% enumeration after a successful application of sampling in 1971, it was decided to use the same sampling procedure in 1976 as in 1971.

Most of the methodology used in the 1971 and 1976 censuses was kept for the 1981 Census, except that the sampling rate was reduced from every third occupied private household to every fifth. Studies done at the time showed that the resulting reduction in data quality (measured in terms of variance) would be tolerable, and would not be significant enough to offset the benefits of reduced cost and response burden, and improved timeliness (see Royce (1983)). Twelve questions were asked on a 100% basis and an additional 34 questions were asked of the sample.

The 1986 Census was the first full mid-decade census. It was decided that only a full census could meet the growing need for local labour market data, a need made more pressing by the occurrence of a major recession (1981-82) since the previous census. However, in order to keep development costs as low as possible, a policy of minimum change was adopted. Unless there were compelling reasons not to do so, 1981 Census questions and data collection and processing procedures were retained. Questions on eight subjects from the 1981 Census were not asked in 1986, while three new questions were added.

---

[3] A collective dwelling is a dwelling of a commercial, institutional or communal nature. Examples include hotels, hospitals, staff residences and work camps.

## B. The Sampling Scheme Used in the 1986 Census

A wealth of information was collected from everyone in Canada on Census Day, 1986. The bulk of the information was acquired on a sample basis. In all self-enumeration areas a 1 in 5 sample of private occupied households was selected to receive a long form (Form 2B), containing all census questions. Nine basic questions on age, sex, marital status, mother tongue, relationship to the household reference person (Person 1), dwelling type and tenure, plus four more dwelling and 19 socio-economic questions were asked. The remaining private dwellings received a short form (Form 2A), containing only the 9 basic census questions.

All dwellings in those areas enumerated by the canvasser method (generally remote areas or Indian Reserves) received the Form 2B. All collective dwellings also received the Form 2B. However, the following persons in collective dwellings were not asked the sample questions:

(a) inmates in correctional and penal institutions or jails;

(b) patients in general hospitals, special care homes and institutions for the elderly, and chronically ill or psychiatric institutions;

(c) children in orphanages and children's homes or young offenders facilities.

Canadians stationed abroad (generally embassy or armed forces personnel) were given a Form 2C, which contained the same questions as the Form 2B except that housing questions were not included. However, questions about the person's usual place of residence in Canada were asked. Information on unoccupied private dwellings was recorded on a Form 2A.

The basic drop-off or delivery procedure required the Census Representative (CR) to pre-plan a route covering all dwellings in his/her enumeration area (EA) and then to visit each dwelling and leave a census questionnaire. The selection of the sample, i.e., the decision as to which type of questionnaire to leave at each occupied dwelling, was facilitated by the Visitation Record (VR), the document in which the CR listed each dwelling in his/her area. This document was printed so that every fifth line was shaded to signify that a Form 2B should be delivered. A random start was implemented by deleting either zero, one, two, three or four lines at the start of the VR according to whether the fifth, fourth, third, second or first dwelling in the EA was to be the first to receive the long form. Thereafter, the dwelling listed on each shaded line automatically received the long form. These procedures were spelled out in the CR's Manual and emphasized in his/her training in order to minimize the risk of any deviation from the specified procedure for selecting the sample.

In sampling terminology, the sample can be described as a stratified systematic sample of private occupied dwellings using a constant 1 in 5 sampling rate in all strata (EAs). As a sample of persons, it can be regarded as a stratified systematic cluster sample with dwellings as clusters. For a more detailed description of the concepts and terminology of sampling, see Stuart (1976), or Cochran (1977).

## C. Processing the Census Sample

Once the CR had obtained the completed questionnaire (Form 2A or 2B) from each dwelling in his/her area, and this work had been approved, the questionnaires were sent to one of seven Regional Processing Sites for manual processing. Complete data for each EA were captured and stored on magnetic tapes. The questionnaires and magnetic tapes were then sent to Head Office Processing in Ottawa. Once there, checks were performed by computer for various inconsistencies in the data which required a manual review of the questionnaire to resolve. After all resulting updates to the data for an EA were completed, the data were reformatted and transferred to Edit and Imputation.

The data were loaded to 10 Edit and Imputation data bases, organized by 2A (100%) and 2B (20%), with 5 regions for each. The 2A data bases contained the basic demographic characteristics for 100% of the population, while the 2B data bases contained the data for the 20% sample questions. The data were processed through a series of customized modules, where all problems of invalid, inconsistent, and missing data were resolved. The 2A data bases were processed first, and a final 2A Canada Retrieval Data Base was created.

Once the 100% data were finalized, the data for the 20% sample questions were processed. Non-response 2B records were dropped from the 2B data bases. A final 2B Canada Retrieval Data Base was created which contained both the 100% and 20% data for sampled households and persons only. The weights created using the 100% data (as described in Chapter III) were placed on this data base.

## III. ESTIMATION FROM THE CENSUS SAMPLE

Any sampling procedure requires an associated estimation procedure for scaling sample data up to the full population level. The choice of an estimation procedure is generally governed by both operational and theoretical constraints. From the operational viewpoint, the procedure must be feasible within the processing system of which it is a part, while from the theoretical viewpoint the procedure should minimize the sampling error of the estimates it produces. In the following two sections, the operational and theoretical considerations relevant to the choice of estimation procedures for the census sample are described.

### A. Operational Considerations

Mathematically, an estimation procedure can be described by an algebraic formula that shows how the value of the estimator for the population is calculated as a function of the observed sample values. In small surveys that collect only one or two characteristics, or in cases where the estimation formula is very simple, it might be possible to calculate the sample estimates by applying the given formula to the sample data for each estimate required. However, in a survey or census in which a wide range of characteristics is collected, or in which the estimation formula is at all complex, the procedure of applying a formula separately for each estimate required is not feasible. In the case of a census for example, every cell of every tabulation based on sample data at every geographic level represents a sample estimate which under this approach would require a separate application of the estimation formula. In addition, the calculation of each estimate separately would not necessarily lead to consistency between the various estimates made from the same census sample.

The approach taken in the census therefore (and in many sample surveys) is to split the estimation procedure into two stages: (a) the calculation of weights (known as the weighting procedure); (b) the summing of weights to produce estimated population counts. Any mathematical complexity is then contained in step (a) which is performed just once, while step (b) is reduced to a simple process of summing weights which takes place at the time a tabulation is retrieved. Also, since the weight attached to each sample unit is the same for whatever tabulation is being retrieved, consistency between different estimates based on sample data is assured.

### B. Theoretical Considerations

For a given sample design and a given estimation procedure, one can, from sampling theory, make a statement about the chances that a certain interval will contain the unknown population value being estimated. The primary criterion in the choice of an estimation procedure is minimization of the width of such intervals so that these statements about the unknown population values are as precise as possible. The usual measure of precision for comparing estimation procedures is known as the standard error. Provided that certain relatively mild conditions are met, intervals of plus or minus two standard errors from the estimate will contain the population value for approximately 95% of all possible samples.

As well as minimizing standard error, a second objective in the choice of estimation procedure for the census sample is to ensure, as far as possible, that sample estimates for basic (i.e., 2A) characteristics are consistent with the corresponding known population values. Fortunately, these two objectives are usually complementary in the sense that sampling error tends to be reduced by ensuring that sample estimates for certain basic characteristics are consistent with the corresponding population figures. While this is true in general, however, forcing sample estimates for basic characteristics to be consistent with corresponding population figures for very small subgroups can have a detrimental effect on the sampling error of estimates for the sample characteristics themselves.

In the absence of any information about the population being sampled other than that collected for sample units, the estimation procedure would be restricted to weighting the sample units inversely to their probabilities of selection (e.g., if all units had a one in 5 chance of selection, then all selected units would receive a weight of 5). In practice, however, one almost always has some supplementary knowledge about the population (e.g., its total size, and possibly its breakdown by a certain variable - perhaps by province). Such information can be used to improve the estimation formula so as to produce estimates with a greater chance of lying close to the unknown population value. In the case of the census sample, a large amount of very detailed information about the population being sampled is available in the form of the basic 100% data at every geographic level. On the one hand, we can take advantage of this population information to improve the estimates made from the census sample; on the other hand, this wealth of information can also be an embarrassment in the sense that it is impossible to make the sample estimates for basic characteristics consistent with all the population information at every geographic level. Differences between sample estimates and population values become visible when a cross-tabulation of a sample variable and a basic variable is produced. The tabulation has to be based on sample data with the result that the marginal totals for the basic variable are sample estimates that can be compared with the corresponding population figures appearing in a different tabulation based on 100% data. They will not necessarily agree exactly.

## C. Developing an Estimation Procedure for the Census Sample

Given that a weight has to be assigned to each unit (person, family or household) in the sample, the simplest procedure would be to give each unit a weight of 5 (because a 1 in 5 sample was selected). Such a procedure would be simple and unbiased[4] and, if nothing but the sample data were known, it might be the optimum procedure. However, although we know that the sample will contain almost exactly one fifth of all households (excluding collective households and those in canvasser areas), one cannot be certain that it will contain exactly one fifth of all persons, or one-fifth of each type of household, or one fifth of all females aged 25-34, and so on. Therefore, this procedure would not ensure consistency even for the most important subgroups of the population. For large subgroups, these fractions should be very close to one fifth, but for smaller subgroups they could differ

[4] "Unbiased" means that the average of the estimates obtained by this procedure, over all possible samples, would equal the true population value.

markedly from one fifth. The next most simple procedure would be to define certain important subgroups (e.g., age-sex groups within province) and, for each subgroup, to count the number of units in the population in the subgroup (N) and the number in the sample (n) and to assign to each sample unit in the subgroup a weight equal to N/n.

For example, if there were 5,000 males aged 20-24 enumerated in Prince Edward Island, and 1,020 of these fell in the sample households, then a weight of 5,000/1,020 = 4.90 would be assigned to each male aged 20-24 in the sample in Prince Edward Island. This would ensure that whenever sex and age in five-year groups were cross-classified against a sample characteristic for Prince Edward Island, the marginal total for the male 20-24 age-sex group would agree with the population total of 5,000. Note that a weight of 5 in this case would result in a sample estimate of 5,100 (1,020 x 5).

This type of estimation procedure is known as "ratio estimation". It can be shown that this procedure can lead to substantial reductions in standard error in many situations. This procedure will ensure consistency between sample estimates and population figures for the chosen subgroups and for combinations of these subgroups. It will not, however, ensure consistency for smaller groups (e.g., counties, or single years of age), nor for groups defined in terms of other basic characteristics (e.g., marital status, mother tongue). One might consider therefore extending this procedure to smaller subgroups defined as the cells in a cross-classification of all relevant basic characteristics. The problem is that, as the subgroup becomes smaller, this procedure becomes unstable (i.e., the standard errors of the estimates produced by this procedure increase). In the limit, the procedure becomes impossible when no sample units happen to fall in a particular subgroup. The challenge, therefore, is to obtain the advantages of ratio estimation without suffering the instabilities of using small subgroups. The solution adopted is to carry out ratio estimation iteratively for two distinct and exhaustive sets of subgroups. This procedure, known as the "raking ratio estimation procedure (RREP)", was used in the 1986 Census and is described in the following section.

## D. The Raking Ratio Estimation Procedure

Instead of just one set of subgroups, two sets of subgroups are defined. One set of subgroups forms the rows of a "weighting matrix" while the other set forms the columns (e.g., for calculating person weights, age-sex-marital status subgroups form the rows of the matrix, while family status-mother tongue subgroups form the columns of the matrix).

Given the appropriate matrix, the RREP proceeds as follows:

(a) Cross-classify the population records into the matrix to give population totals in each row and column.

(b) Cross-classify the sample records into the same matrix to give sample counts in each cell and sample totals in each row and column.

(c) If necessary, collapse the rows and columns of the matrix to meet certain size constraints (see below).

(d) Assign an initial weight of 5 to each sample record.

(e) For each column, compare the estimated column total using this initial weight to the known column population total. Eliminate any discrepancies at the column level by multiplying the initial weights by the ratio of the column population total to the estimated column total. These revised weights are called the first iteration weights.

(f) For each row, compare the estimated row total using these first iteration weights to the known row population total. Eliminate any discrepancies at the row level by multiplying the first iteration weights by the ratio of the row population total to the estimated row total (this will destroy the exact agreement for columns).

(g) Continue this process of eliminating discrepancies in the column and row estimates until any remaining discrepancies are negligible (when the process is said to have "converged") or to a maximum of 80 iterations.

The procedure stops on rows so that row totals are exactly consistent and column totals are almost exactly consistent. The important feature of this procedure is that the size constraints (to avoid instability in the estimators) apply only to the row and column totals and not to the individual cells of the matrix (some of which could even be empty). For more details on the RREP, see Brackstone and Rao (1979). The RREP is based on a procedure which has come to be known as "Iterative Proportional Fitting". This procedure was first proposed in Deming and Stephan (1940).

There are two parameters in the RREP which are crucial to the question of consistency between sample estimates for basic characteristics and the corresponding population figures. The first is the choice of the geographic area or weighting area (WA) within which the above procedure is applied. Steps (a) to (g) described above are applied independently within each WA. The second is the choice of the subgroups to define the rows and columns of the weighting matrix.

The WA is the geographic area for which almost exact agreement is ensured for total counts of persons and households and for those subgroups defined by the rows and columns of the weighting matrix. From the point of view of consistency for small areas, the smaller the WA the better. However, the smaller the WA the less detail is possible in the rows and columns of the weighting matrix (because of minimum size limits on these rows and columns). The compromise that was adopted for the 1986 Census was the following:

(a) a WA should contain between 2,000 and 7,000 persons (100% count);

(b) WA boundaries must respect the boundaries of census divisions (CDs), and as far as possible, of census subdivisions (CSDs), census tracts (CTs), and federal electoral districts (FEDs);

(c) WAs should be made up of whole EAs and should generally be connected (i.e., no "holes").

There are two criteria for choosing subgroups to use in the weighting matrix. First, correlation between the variables defining the subgroups and the sample characteristics is important in minimizing the sampling error of the sample estimates. Secondly, the need to

ensure consistency for certain important subgroups will influence the choice of rows and columns. These two criteria are often (but not necessarily) complementary. Because of the size constraints on rows and columns the matrix cannot be too detailed. Two different matrices were used in 1986, one for person and family weights, the other for household weights. The two matrices are shown in the Appendix.

Associated with each matrix was a collapsing strategy that defined how rows and columns were to be combined in any WA in which a row or column met one of the following criteria:

(a) the row or column population total was less than 35;

(b) the row or column sample count was zero;

(c) the ratio of the population count to the sample count for the row or column was not in the range 3.0 to 19.9.

The choice of collapsing strategies was designed to preserve subgroups wherever possible. For example, when necessary, the "Rented Other" column in the household matrix was collapsed with the "Rented Apartment" column, not with one of the "Owned" columns, so that the "Rented" subgroup was preserved. Collapsing was carried out independently within each WA and ended as soon as all row and column population and ratio constraints were satisfied. The matrices given in the Appendix have single and double lines drawn on them to divide the rows and columns into groups. Collapsing took place initially within groups separated by single lines (where they exist). Then, if necessary, collapsing took place across single lines within double lines. Only on rare occasions did collapsing take place across double lines. This occurred when all the rows or columns between two double lines had been collapsed together into one row or column for which the sample count was still zero but the population count was greater than zero.

The RREP resulted in final weights that were the same for all units in the same cell of the collapsed matrix but which differed from cell to cell. These final weights were then added to the record of each sample unit on the data base. Each person in the sample received the weight calculated for the cell of the person and family matrix in which he/she fell; each household in the sample received the weight from the appropriate cell of the household matrix; each census family in the sample received the personal weight of the husband or lone parent in the family. The weight of the husband of the economic family reference person or, if the husband was not present, then the economic family reference person was used as the weight for the economic family. Persons, households, and families in those sectors of the population enumerated on a 100% basis automatically received a weight equal to one.

Operationally, the RREP was almost fully automated. Weighting areas were formed using a computer program that takes into account EA population, geographic co-ordinates of EA centroids, and the geostatistical area (CDs, CSDs, etc.) in which the EA is located. This program provided a listing of the WAs thus formed and allowed changes to be made manually if appropriate. This facility for manual adjustment was used in a small number of cases. Once the WAs had been fixed, automated procedures were used for the cross-classification of data, the collapsing of rows and columns, the calculation of weights, and the assignment of these weights to records on the data base.

## IV. THE SAMPLING AND WEIGHTING EVALUATION PROGRAM

The sampling and weighting evaluation program was designed to determine the effect of sampling and weighting on the quality of census sample data. To this end, five studies were carried out to measure the quality of the census sample data and estimates and to provide information relevant to the planning of future censuses. These studies were:

(a) an examination of sampling bias;

(b) an evaluation of the formation of weighting areas;

(c) an evaluation of the weighting procedures;

(d) an evaluation of sample estimate and population count consistency;

(e) a study to produce estimates of variance for various 20% sample characteristics.

In the remainder of this chapter, these five studies are briefly described. Chapters V through VIII present the results of these studies.

### A. Sampling Bias Study

Bias can be introduced into responses to any survey from a number of sources. The objective of this study was to determine if responses to basic questions on Forms 2B were biased in any way and to identify, if possible, the causes of any observed bias.

### B. Evaluation of Weighting Area Formation

The objective of this study was to measure the degree to which WAs met the criteria laid down for their formation (see Chapter III, Section D). All WAs in Canada were analyzed to determine how well they respected the size constraints and the boundaries of various types of geographic areas. Causes of violations of size criteria were investigated.

### C. Evaluation of Weighting Procedures

The objective of this study was to evaluate the performance of the RREP. The level of agreement between the sample estimates and population counts for the rows and columns of the cross-classification matrices of all WAs in Canada was examined. The amount of collapsing of rows and columns, the degree of convergence of the RREP, and the variability in sampling fractions and population sizes among rows and columns were studied to explain observed inconsistencies.

### D. Sample Estimate and Population Count Consistency Study

This study examined the level of agreement (consistency) between sample estimates and population counts for a wide variety of basic characteristics, not just those used to define the rows and columns of the cross-classification matrices. The consistency was studied for

various types of geographic areas other than WAs, whose boundaries were not always respected by WA boundaries. In addition, the consistency for the individual cells of the cross-classification matrices were examined. A separate study was done on the consistency of the characteristic "mother tongue".

**E. Sampling Variance Study**

The "variance" of an estimate is a measure of its precision. Estimates of variance for estimates using simple weights of 5 and assuming simple random sampling are relatively inexpensive to calculate. However, estimates of variance for raking ratio estimates taking into account the sample design used are very expensive to calculate. The objective of this study was to develop an inexpensive method of producing these estimates of variance. This was done by calculating "adjustment factors", which are the ratios of the estimates of the standard errors (the square roots of the variances) for raking ratio estimates to the simple estimates of the standard errors. An estimate of the standard error of a raking ratio estimate for any characteristic in any geographic area can then be obtained by multiplying the simple estimate of the standard error by the appropriate adjustment factor.

## V. SAMPLING BIAS

Estimates based on a sample survey are subject to sampling errors. One type of sampling error arises from the variability in the population. This variability means that different samples will produce different estimates, none of which will necessarily equal the true population value. The estimates will equal the true population value on average, however, provided that there is no bias in the sample creating a tendency to over-estimate or under-estimate. Unfortunately, bias is often difficult to eliminate completely. In the Census of Population, bias can be introduced into the responses from a variety of sources. These include coverage errors, non-response bias, response bias (e.g., respondents answering differently on the Form 2B than on the 2A), CR errors (e.g., not selecting the sample according to specifications), processing errors, and so on.

The purpose of the Sampling Bias Study was to search for bias in the responses to the basic questions on Forms 2B. Sample estimates for a wide variety of basic characteristics were compared to the population counts for all 260 sampled census divisions (CDs) in Canada. The sample estimates were produced by multiplying the sample counts at the EA level by simple weights equal to the inverse of the EA sampling fraction (approximately 5) and then summing to the CD level[5]. Plots of the differences between the sample estimates and the population counts for each CD were produced separately for each characteristic. This was done to see if patterns existed which would indicate definite tendencies for estimates to be too low (biased downward) or too high (biased upward). In addition, tests were done to determine if the differences between the sample estimates and population counts were statistically significant.

The pattern of differences exhibited in the plots indicated that some degree of bias was indeed present in the sample for most characteristics. Furthermore, the average difference between the sample estimates and the population counts, over all CDs, was found to be statistically significant (at the 5% level[6]) for most of the characteristics (i.e., the differences cannot be explained by sampling variability). Table 1 shows the differences (in absolute and percentage terms) between the sample estimates and the population counts at the Canada level (averaged over all CDs) for all characteristics studied. In most cases the bias was less than 1%. Also given is the percentage of CDs in which each characteristic was over-represented. A percentage less than 50 means that the characteristic was under-represented in a majority of CDs. The conclusions drawn from the analysis of the differences are given in the following paragraphs.

The sizes of the households in the 2B sample were larger on average than for the total population. There was a definite tendency for the following groups of people to be over-represented in the sample: females, age groups 0-5, 6-14, 35-44, and 45-54, and census family persons, in particular married persons and census family children.

---

[5] These simple estimates were used instead of the raking ratio estimates because the RREP reduces the sampling bias by forcing estimates of basic characteristics to equal population counts.

[6] This means that there was at most a 5% chance of obtaining such large differences in the absence of bias.

**Table 1. Sample Estimate (Simple Weights) Minus Population Count at Canada Level (Sampled EAs Only) and Percentage of CDs in which Characteristic was Over- Represented**

| Characteristic | Sample Estimate Minus Population Count | | | Percentage of Over-Represented CDs |
|---|---|---|---|---|
| **Person Characteristics** | | | | |
| Males | 1,557 | * | (+0.01%) | 54 |
| Females | 22,017 | * | (+0.18%) | 73 |
| Total Person Population | 23,574 | * | (+0.10%) | 71 |
| Age 0-5 | 9,185 | * | (+0.44%) | 69 |
| Age 6-14 | 17,787 | * | (+0.57%) | 75 |
| Age 15-24 | - 8,385 | * | (- 0.21%) | 43 |
| Age 25-34 | - 1,332 | * | (- 0.03%) | 56 |
| Age 35-44 | 5,661 | * | (+0.16%) | 60 |
| Age 45-54 | 2,844 | * | (+0.11%) | 57 |
| Age 55-64 | 2,360 | | (+0.10%) | 49 |
| Age 65 and Over | - 4,546 | * | (- 0.19%) | 33 |
| Single Persons | - 462 | * | (- 0.00%) | 55 |
| Married Persons | 37,195 | * | (+0.32%) | 83 |
| Widowed Persons | - 5,405 | * | (- 0.49%) | 29 |
| Divorced Persons | - 3,937 | * | (- 0.59%) | 41 |
| Separated Persons | - 3,817 | * | (- 0.77%) | 39 |
| **Family Characteristics** | | | | |
| Total # of Census Families | 20,056 | * | (+0.30%) | 85 |
| Husband-Wife Census Families | 20,250 | * | (+0.35%) | 85 |
| Lone Parent Census Families | - 194 | | (- 0.02%) | 45 |
| Census Family Children | 30,418 | * | (+0.36%) | 76 |
| People in Census Families | 70,724 | * | (+0.34%) | 84 |
| People Not in Census Families | - 47,150 | * | (- 1.35%) | 9 |
| **Household and Dwelling Characteristics** | | | | |
| Owned Dwellings | 5,995 | * | (+0.11%) | 60 |
| Rented Dwellings | - 5,995 | * | (- 0.18%) | 40 |
| Single Detached Dwellings | 6,792 | * | (+0.13%) | 68 |
| Apts With Less Than 5 Storeys | - 5,315 | * | (- 0.31%) | 31 |
| Apts With 5 or More Storeys | - 209 | | (- 0.03%) | 43 |
| Movable Dwellings | 90 | | (+0.08%) | 51 |
| All Other Types of Dwellings | - 1,358 | * | (- 0.12%) | 42 |
| One Person Households | - 13,248 | * | (- 0.70%) | 22 |
| Two Person Households | 4,082 | | (+0.15%) | 49 |
| Three Person Households | 2,413 | * | (+0.15%) | 57 |
| Four or Five Person Households | 8,459 | * | (+0.36%) | 72 |
| Six or More Person Households | - 1,706 | | (- 0.52%) | 47 |
| Non Census Family Households | - 21,249 | * | (- 0.92%) | 14 |
| One Census Family Households | 22,349 | * | (+0.35%) | 87 |
| Multiple Census Family Hhlds | - 1,100 | * | (- 1.17%) | 39 |
| Hhld Maintainers Aged < 25 | - 3,739 | * | (- 0.71%) | 40 |
| Hhld Maintainers Aged 25-34 | 1,673 | * | (+0.08%) | 59 |
| Hhld Maintainers Aged 35-44 | 3,385 | * | (+0.17%) | 59 |
| Hhld Maintainers Aged 45-64 | 2,302 | | (+0.09%) | 57 |
| Hhld Maintainers Aged > 64 | - 3,621 | * | (- 0.23%) | 30 |
| Male Households Maintainers | 1,025 | | (+0.02%) | 52 |
| Female Household Maintainers | - 1,025 | | (- 0.04%) | 48 |

* These differences were found to be statistically significant at the 5% level.

The following groups of people were under-represented in the sample: age groups 15-24 and greater than 64, widowed, divorced, and separated persons, and non-census family persons. The under-representation of these person characteristics is particularly significant given that on average there were more people in sampled dwellings than non-sampled dwellings.

In terms of household characteristics, there was a tendency for owned dwellings and single detached dwellings to be over-represented in the sample, while rented dwellings and apartments with less than 5 storeys tended to be under-represented. There was a tendency for one census family households to be over-represented, while non-census family households and, to a lesser extent, multiple census family households were under-represented. Consistent with this, there was a tendency for four or five person households to be over-represented while one person households were under-represented. Household maintainers aged 25-34 and 35-44 were over-represented, while those aged less than 25 and greater than 64 were under-represented.

As mentioned above, there are many possible explanations for the observed differences between the sample estimates based on simple weights and the population counts. One possibility arises from the fact that there were 67,884 (0.8% of the total) complete non-response households in the 1986 Census. These were either households which completely refused to answer the questions or for which the CR was unable to get any information (usually because the members of the household were absent during the census-taking period or had moved on or after census day without responding). The percentage of sampled households which were non-response was more than twice as high as the percentage of non-sampled households. It is possible that non-response households had different characteristics in general than households which responded (e.g., they could have been smaller). If so, then the sample data would have been disproportionately affected. Non-response bias would have been introduced into both the sample and 100% data, and sample estimate and population count discrepancies would have been created as a result of the bias being larger for sampled households.

During data processing, complete non-response sampled households were removed from the sample (so that they became non-sampled households) and the responses to the basic questions only were imputed. Therefore, if the imputation system had a tendency to impute certain types of households more often than others, this would have caused sample estimate and population count discrepancies as well, since only non-sampled households would have been affected. In this case, the 100% data would have been biased, but not the sample data. When non-response households were removed from the study, there was found to be some reduction in the amount of bias observed for most characteristics. However, the remaining bias was still statistically significant at the 5% level for 33 of the 44 characteristics studied (the bias was significant for 35 of the characteristics when non-response households were included). The impact of imputation of complete non-response households, therefore, was not large enough to explain all of the observed bias.

Other possible sources of bias were also studied. During data processing, sample households were also converted to non-sample households in the case where the basic questions were answered by the respondent, but all of the sample questions were left unanswered. This would create sample estimate and population count discrepancies if certain types of households had a greater tendency than others not to respond to the sample questions.

Partial non-response refers to the situation where some, but not all, of the questions are left unanswered by the respondent. Answers to these questions were imputed by the system. A higher rate of partial non-response on the Form 2B than on the Form 2A, or vice-versa, would result in sample estimate and population count discrepancies if certain types of households or people had a greater tendency than others to not respond to certain questions, or if the imputation system had a tendency to impute certain responses at an inappropriate frequency.

When the impact due to these last two factors was removed from the data, there was a further reduction in the amount of observed bias (on top of that resulting from the removal of complete non-response households) for most characteristics. The remaining bias was, however, still statistically significant for 32 of the 44 characteristics at the 5% level. Consequently, although these factors did seem to contribute to the bias, much of it remains unexplained.

Another possible source of bias was the fact that persons living in sampled households were missed at a higher rate than persons living in non-sampled households. It is also known that the characteristics of missed persons differ from those of enumerated persons. Bias would thus be introduced into both the 100% and sample data, but because there is more undercoverage of persons in sampled households, sample and population count discrepancies would be created. For more information on coverage in the 1986 Census, see the User's Guide to the Quality of 1986 Census Data: Coverage.

For more information on the Sampling Bias Study, see Rathwell (1990).

## VI. EVALUATION OF WEIGHTING PROCEDURES

### A. Weighting Area (WA) Formation

The first stage of the weighting procedures was the formation of WAs. The objectives of WA formation were to create WAs large enough for the RREP to work well (a population of at least 2,000), but small enough to respect the boundaries of as many census subdivisions (CSDs), census tracts (CTs) and federal electoral districts (FEDs) as possible. As well, WAs had to respect the boundaries of all census divisions (CDs). The sampled EAs were formed into 5,341 WAs[7] with an average population (excluding persons in collective dwellings) of 4,558. Of the 5,341 WAs, 5,229 (98%) fell within the population range of 3,000-7,000. Of the remaining 112 WAs, 107 were in the range 2,001-2,999, 93% of these being in the range 2,501-2,999. The remaining five WAs were in the range 1-2,000. These had been created specially to correspond to EAs with extreme sampling fractions (close to 0% or 100%).

The extent to which WAs respected the boundaries of various geographic areas was examined separately for CTs, CSDs in census-tracted areas, CSDs in non census-tracted areas and FEDs. Since CD boundaries were always respected, no study was necessary for them. Only the sampled portion of geographic areas were considered in verifying the respect for boundaries. Geographic areas which did not contain any sampled EAs were excluded from the study.

Table 2 shows how well the boundaries of CTs, CSDs and FEDs were respected by WAs. The first column shows the percentage of geographic areas which contained only entire WAs. The second column shows the percentage of geographic areas which were too small to form entire WAs, but were completely contained within one WA. The third column shows the percentage which contained parts of different WAs.

---

[7] An additional six WAs were formed by the automated system, but since they contained no sampled EAs they were not used in the RREP.

**Table 2. Extent to which Weighting Areas Respected Various Geographic Boundaries**

| Geographic Areas | Contained Only Entire WAs | Contained Entirely Within One WA | Contained Parts of Different WAs |
|---|---|---|---|
| census divisions | 100% | 0% | 0% |
| census tracts | 56% | 32% | 12% |
| census subdivisions in census-tracted areas | 57% | 32% | 11% |
| census subdivisions in non census-tracted areas | 8% | 85% | 7% |
| federal electoral districts | 15% | 0% | 85% |

Since the RREP is performed independently within WAs, agreement between sample estimates and population counts is ensured only for those geographic areas which contain only entire WAs. Agreement is not ensured for geographic areas which are completely contained within one WA or which contain parts of different WAs.

The WAs respected the boundaries of CTs and CSDs in census-tracted areas almost equally well. The small size of most CSDs in non census-tracted areas resulted in most of them (85%) being contained within one WA.

Just 15% of FEDs contained only whole WAs. Also, because FEDs are considerably larger than WAs, none were completely contained within a part of one WA. A majority of FEDs (80%), contained between 5 and 30 whole WAs and 1 to 10 partial WAs. Their boundaries were thus not well respected in the formation of WAs. Since FED boundaries are completely unrelated to CD, CT, and CSD boundaries, they could not be respected better without the risk of further violating these boundaries.

For more information on this study, see Daoust (1987).

## B. Evaluation of the Raking Ratio Estimation Procedure

One of the aims of the weighting procedure is to minimize the discrepancies between population counts defined by the rows and columns of the weighting cross-classification matrices and the corresponding sample estimates. These discrepancies are the result of sampling variability and bias (see Chapter V). Even after the weighting procedure is completed, however, some discrepancies may remain. One of the main causes of such

discrepancies is the collapsing of rows and columns together before the weighting procedure begins. This is done in order to satisfy size constraints which help to ensure that the weighting procedure works well (see Chapter III, Section D). Discrepancies are measured by the difference between the sample estimate and the population count, expressed as a percentage of the population count, i.e.,

$$\text{discrepancy} = \frac{\text{sample estimate - population count}}{\text{population count}} \times 100$$

Discrepancies can become large if a small row and a large row with quite different sampling fractions[8] are collapsed together. For example, suppose a row with a population of 30 and a sample of 5 (i.e., the percentage sampled is 16.7%) is collapsed with a row that has a population of 90 and a sample of 19 ( i.e., the percentage sampled is 21.1%). Combining them produces a collapsed row with a population of 120 and a sample of 24. The weight of the combined row could be defined as 120/24 = 5, which is the ratio of the population count to the sample count (this is a simplification, of course, to what is actually done by the RREP). Using this weight, the estimate in the smaller row would be 5 x 5 = 25. This would generate a discrepancy of (25 - 30)/30 x 100 = -16.7%. Applying the weight to the larger row would result in an estimate of 19 x 5 = 95. The discrepancy would be (95 - 90)/90 x 100 = 5.6%. Thus the result is a large discrepancy for the small row and a somewhat smaller discrepancy for the large row.

Discrepancies were calculated for each row and column of both the person and household matrices, for each major region of the country (east, Quebec, Ontario and west, including the Territories). The discrepancies for all of the rows of the household matrix were less than ± 4% for all regions of Canada. For example, Figure 1 gives the discrepancies for Canada and east. The vertical axis represents the discrepancy in percentage terms as given by the above formula. The numbers on the horizontal axis correspond to the row numbers of the household matrix as given in the Appendix. The vertical lines indicate where the double lines are on the matrix as shown in the Appendix, across which collapsing rarely occurred (see Chapter III, Section D). In many cases the large discrepancies were caused by collapsing small rows with large rows having quite different sampling fractions. For example, from Figure 1 it can be seen that row 22 had a discrepancy of approximately - 2% for the east region. Row 22 contains one person non-family households with a male household maintainer aged 65 or greater. Figure 2 gives the percentage of the population (in sampled EAs only) that belonged to each row. It can be seen that only approximately 2% of the households in the east fell in this row. Figure 3 gives the percentage of WAs for which the rows were collapsed. Row 22 was collapsed nearly 80% of the time in the east. Most of the time it was collapsed with row 23 alone (one person non-family households with a female household maintainer aged 65 or greater), which had 3 times the population of row 22. Furthermore, as seen in Figure 4, the sampling fraction of row 22 was less than 18% in the

[8] Differences in sampling fractions are caused by such things as sampling variability, corrections for non-response, sampling bias or response bias.

Figure 1. Discrepancy By Rows of Household Weighting Matrix

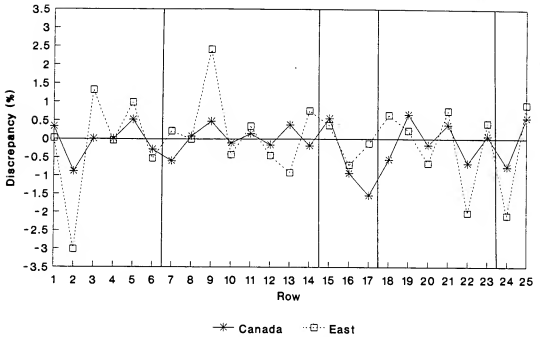



Figure 2. Percentage of Population by Rows of Household Weighting Matrix

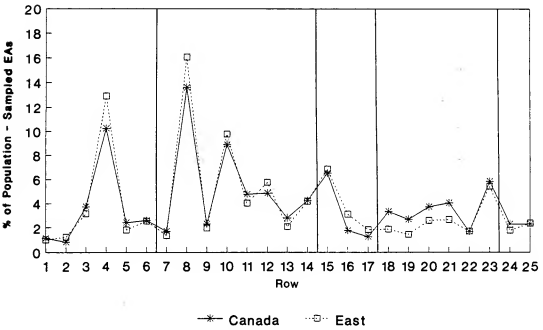


Figure 3. Percentage of Collapsing By Rows of Household Weighting Matrix

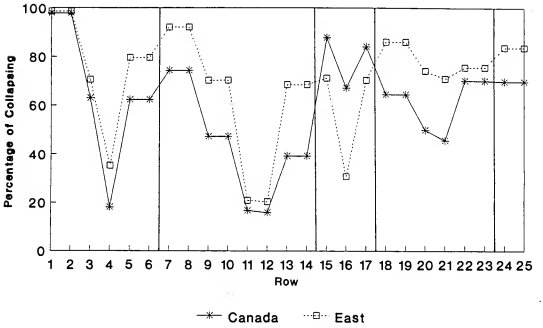
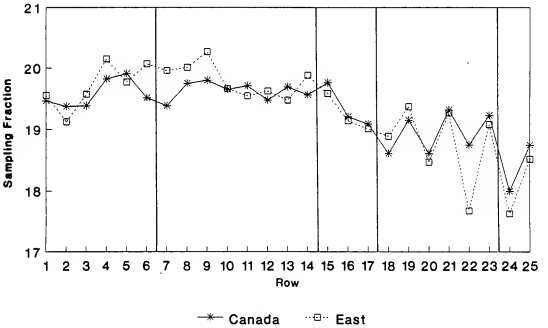



Figure 4. Sampling Fraction By Rows of Household Weighting Matrix



east while in row 23 it was more than 19%. Thus it is not surprising that row 22 had a - 2% discrepancy in the east region.

There were far fewer columns than rows (4 compared to 25) in the household matrix. As a result, relatively little collapsing was necessary. In addition, the sampling fractions were fairly similar for each column. Consequently, the columns of the household matrix had very small discrepancies at the regional level (within 0.13% for all the columns in each region).

All the rows of the person matrix had discrepancies of less than ± 0.1% in all regions except rows 6, 7, 8, 9, 19, 20, 21 and 22. These rows all had discrepancies of less than ± 1% in all regions except for row 7 in Ontario for which the discrepancy was less than 3%. These discrepancies were all caused by small rows being collapsed with large rows with different sampling fractions. Row 6 was frequently collapsed with row 7, 8 with 9, 19 with 20 and 21 with 22. Row 6 is much larger than row 7, 9 is much larger than 8, 19 is much larger than 20 and 22 is much larger than 21.

The "Other" mother tongue columns of the person matrix tended to be under-estimated in the east (negative discrepancies of as much as 9%) and to a lesser extent in Quebec. French mother tongue columns tended to be over-estimated in the west (positive discrepancies of as much as 6%) and to a lesser extent in Ontario. There was also a slight tendency for English to be over-estimated in Quebec (positive discrepancies of less than 2%). These patterns were the result of the English, French and Other mother tongue columns usually all being collapsed together in the east and in Quebec, while French and Other tended to be collapsed together in the west and in Ontario. The sampling fraction for French mother tongue tended to be the highest, followed by English and then Other. The low sampling fraction for Other was partly the result of respondents tending to give multiple responses (which are included in Other) more frequently on the Form 2A than on the 2B (see Chapter VII).

In addition to collapsing, discrepancies between sample estimates and population counts can occur when the RREP fails to converge. A maximum of 80 iterations was allowed for the RREP to converge. This was sufficient for all household matrices, in fact 98% of them required less than 20 iterations. However, the average number of iterations for the person matrices was much higher. In fact, 385 (7.2%) had not yet converged at the end of 80 iterations. This only affected the columns, since the RREP always ends on rows, so that collapsed row discrepancies are always zero. Exact convergence is not required for the RREP to end, and in fact is rarely achieved. However, the discrepancies for the columns can be expected to be larger if the RREP does not reach the level of convergence required at the end of 80 iterations. The lack of convergence may be caused by inconsistent row and column constraints. Inconsistent constraints can arise when, after collapsing of rows and columns has occurred, there is a block of cells in the matrix for which there are no in-sample units, but some in the population. This can create a situation where it is impossible to make the sample estimates equal the population counts for both the rows and the columns simultaneously. The use of age to define both the rows and columns of the person matrix did cause inconsistent constraints in at least one matrix. This resulted in an over-estimation of census family children aged 0 to 14.

For more information on this study, see Daoust and Bankier (1989).

## VII. SAMPLE ESTIMATE AND POPULATION COUNT CONSISTENCY

Size constraints on the rows and columns of the cross-classification matrices, required for the RREP to work well, limited the number of rows and columns the matrices could have. Consequently, many important characteristics were grouped together when the rows and columns were formed. As a result, the level of agreement (consistency) between sample estimates and population counts for these characteristics was reduced. Furthermore, many geographic areas of interest do not always consist of complete WAs (see Chapter VI, Section A). Consequently, in these areas the consistency for all characteristics depends on how close the areas come to consisting of complete WAs.

The consistency study examined the discrepancies between sample estimates and population counts (expressed as percentages of the population counts) for the same basic characteristics as the Sampling Bias Study for the following geographic areas:

(a) census divisions;

(b) census subdivisions;

(c) census tracts and provincial census tracts;

(d) enumeration areas.

In addition, consistency was examined for:

(e) the cells of the weighting matrices;

(f) the mother tongue characteristic.

As in Section 6.2, the discrepancies between sample estimates and population counts were calculated as:

$$\text{discrepancy} = \frac{\text{sample estimate - population count}}{\text{population count}} \times 100$$

### A. Census Divisions (CDs)

The percentiles in Table 3 summarize the level of consistency for all 260 sampled CDs in Canada for a wide variety of basic characteristics. For each characteristic, N% of the CDs had discrepancies that were less than the Nth percentile while 100 - N% of the CDs had discrepancies that were greater than the Nth percentile. Thus, the discrepancy was between the 10th and 90th percentiles for 80% of the CDs, was between the 25th and 75th percentiles for 50% of the CDs, etc. For example, the discrepancy for age 0-5 was between -1.62% and 1.58% for 80% of the CDs.

**Table 3. Percentiles of Sample Estimate and Population Count Discrepancies (as a Percentage of the Population Count) for CDs and Percentage of Improved CDs**

| Characteristic | Percentiles of Discrepancies | | | | | Percentage of CDs for which Raking Ratio Improved Over Simple Estimates |
|---|---|---|---|---|---|---|
| | 10th | 25th | 50th | 75th | 90th | |
| **Person Characteristics** | | | | | | |
| Males | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 100 |
| Females | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 100 |
| Total Person Population | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 100 |
| Age 0-5 | - 1.62 | - 0.57 | 0.06 | 0.93 | 1.58 | 81 |
| Age 6-14 | - 1.08 | - 0.62 | - 0.05 | 0.36 | 1.02 | 88 |
| Age 15-24 | - 0.46 | - 0.20 | 0.00 | 0.16 | 0.41 | 94 |
| Age 25-34 | - 1.23 | - 0.48 | - 0.01 | 0.39 | 0.83 | 82 |
| Age 35-44 | - 0.90 | - 0.44 | 0.03 | 0.51 | 1.05 | 86 |
| Age 45-54 | - 1.00 | - 0.51 | - 0.03 | 0.47 | 0.91 | 89 |
| Age 55-64 | - 1.06 | - 0.40 | 0.09 | 0.52 | 1.23 | 88 |
| Age 65 and Over | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 100 |
| Single Persons | - 0.28 | - 0.11 | 0.01 | 0.13 | 0.28 | 94 |
| Married Persons | - 0.33 | - 0.17 | - 0.07 | 0.04 | 0.18 | 90 |
| Widowed Persons | - 4.49 | - 2.01 | - 0.07 | 2.14 | 5.26 | 63 |
| Divorced Persons | - 8.45 | - 3.64 | 0.57 | 4.77 | 9.30 | 57 |
| Separated Persons | - 10.60 | - 4.48 | - 0.22 | 5.06 | 10.51 | 60 |
| **Family Characteristics** | | | | | | |
| Total # of Census Families | - 0.13 | - 0.05 | - 0.01 | 0.03 | 0.08 | 100 |
| Husband-Wife Census Families | - 0.13 | - 0.05 | - 0.01 | 0.04 | 0.09 | 100 |
| Lone Parent Census Families | - 0.13 | - 0.07 | - 0.04 | 0.00 | 0.04 | 100 |
| Census Family Children | - 0.08 | - 0.01 | 0.05 | 0.10 | 0.20 | 100 |
| People in Census Families | - 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 100 |
| People Not in Census Families | - 0.12 | - 0.07 | - 0.03 | - 0.00 | 0.03 | 100 |
| **Household and Dwelling Characteristics** | | | | | | |
| Owned Dwellings | - 0.00 | - 0.00 | 0.00 | 0.00 | 0.00 | 100 |
| Rented Dwellings | - 0.00 | - 0.00 | 0.00 | 0.00 | 0.00 | 100 |
| Single Detached Dwellings | - 0.53 | - 0.21 | 0.03 | 0.22 | 0.48 | 74 |
| Apts With Less Than 5 Storeys | - 3.95 | - 1.97 | - 0.55 | 0.37 | 1.77 | 77 |
| Apts With 5 or More Storeys | - 100.00 | - 34.78 | - 0.77 | 2.30 | 16.15 | 33 |
| Movable Dwellings | - 14.43 | - 2.93 | 1.30 | 7.27 | 16.97 | 53 |
| All Other Types of Dwellings | - 6.11 | - 1.76 | 0.26 | 2.59 | 5.27 | 57 |
| One Person Households | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 100 |
| Two Person Households | - 1.65 | - 0.80 | 0.22 | 0.84 | 2.14 | 73 |
| Three Person Households | - 4.36 | - 1.86 | - 0.10 | 1.82 | 3.93 | 62 |
| Four or Five Person Households | - 2.73 | - 1.11 | 0.06 | 1.23 | 2.62 | 63 |
| Six or More Person Households | - 11.43 | - 6.19 | - 1.61 | 2.23 | 7.10 | 52 |
| Non Census Family Households | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 100 |
| One Census Family Households | - 0.23 | - 0.05 | 0.04 | 0.17 | 0.30 | 95 |
| Multiple Census Family Hhlds | - 40.67 | - 18.69 | - 4.17 | 6.04 | 28.75 | 47 |
| Hhld Maintainers Aged < 25 | - 8.07 | - 4.12 | - 0.33 | 2.45 | 6.63 | 64 |
| Hhld Maintainers Aged 25-34 | - 1.76 | - 0.73 | 0.12 | 0.84 | 1.59 | 75 |
| Hhld Maintainers Aged 35-44 | - 1.99 | - 0.81 | 0.04 | 0.94 | 1.75 | 75 |
| Hhld Maintainers Aged 45-64 | - 1.33 | - 0.56 | 0.11 | 0.66 | 1.27 | 79 |
| Hhld Maintainers Aged > 64 | - 1.44 | - 0.61 | - 0.02 | 0.38 | 1.05 | 87 |
| Male Households Maintainers | - 0.75 | - 0.38 | - 0.09 | 0.18 | 0.41 | 73 |
| Female Household Maintainers | - 1.57 | - 0.56 | 0.24 | 1.16 | 2.64 | 73 |

All CDs consist uniquely of complete WAs. Thus the characteristics which were represented by a row or column in the matrix which was rarely or never collapsed had nearly perfect consistency at the CD level[9]. These characteristics were: sex, age 65 and over, all census family characteristics, owned dwellings, rented dwellings, one person households and non census family households. The level of consistency for the remaining characteristics was not perfect but was still quite good, except for those characteristics which represent only a small percentage of the population in most CDs, such as apartments with 5 or more storeys and multiple census family households. Plots (not shown in this report) of the discrepancies against the population counts showed that, in general, the consistency improved as the population count for the CD increased, for all characteristics.

The final column of Table 3 gives the percentage of CDs for which the raking ratio estimate was closer to the population count than the estimate using a simple weight of approximately 5[10]. The raking ratio estimate was better in a majority of CDs for all characteristics except apartments with 5 or more storeys and multiple census family households.

**B. Census Subdivisions (CSDs)**

Table 4 summarizes the level of consistency between sample estimates and population counts for all sampled CSDs in Canada with a population count[11] greater than 50. It covers the same characteristics as Table 3. CSDs do not always consist uniquely of complete WAs. They are also much smaller on average than CDs. Consequently, the consistency was not as good for CSDs as for CDs. The raking ratio estimates were better than the estimates using simple weights for the majority of CSDs for almost all characteristics. In general, as with CDs, the consistency improved as the population count for the CSD increased, for all characteristics.

---

[9] Even for characteristics with perfect consistency, tabulations of basic characteristics based on sample data will not agree exactly with tabulations of the same characteristics based on 100% data. This is because those residents of collective dwellings which were not asked the sample questions (see Chapter II, Section B) are included in tabulations based on 100% data, but are excluded from tabulations based on sample data.

[10] The simple weight (referred to here and elsewhere in this chapter) for each unit (person or dwelling) was actually equal to the inverse of the household sampling fraction for the EA in which the unit was located.

[11] The population count here refers to that of the characteristic. For example, the level of consistency for age 0-5 is summarized for all CSDs in which there were more than 50 people in the age group 0-5. The same definition applies to tables 5, 6, and 7.

**Table 4. Percentiles of Sample Estimate and Population Count Discrepancies (as a Percentage of the Population Count) for CSDs and Percentage of Improved CSDs**

| Characteristic | Percentiles of Discrepancies | | | | | Percentage of CSDs for which Raking Ratio Improved Over Simple Estimates |
|---|---|---|---|---|---|---|
| | 10th | 25th | 50th | 75th | 90th | |
| **Person Characteristics** | | | | | | |
| Males | - 9.15 | - 3.46 | 0.00 | 3.26 | 9.75 | 60 |
| Females | - 9.38 | - 3.44 | 0.00 | 3.34 | 8.97 | 60 |
| Total Person Population | - 7.56 | - 2.71 | 0.00 | 2.59 | 7.93 | 58 |
| Age 0-5 | - 20.36 | - 7.44 | 0.05 | 7.28 | 19.33 | 65 |
| Age 6-14 | - 20.42 | - 7.16 | - 0.16 | 7.14 | 19.61 | 65 |
| Age 15-24 | - 20.30 | - 7.40 | 0.00 | 6.23 | 19.22 | 66 |
| Age 25-34 | - 17.54 | - 6.53 | - 0.05 | 6.29 | 17.55 | 62 |
| Age 35-44 | - 17.96 | - 6.29 | - 0.10 | 6.63 | 18.49 | 64 |
| Age 45-54 | - 19.38 | - 7.09 | - 0.11 | 6.47 | 19.73 | 65 |
| Age 55-64 | - 20.51 | - 7.11 | 0.08 | 7.53 | 19.78 | 65 |
| Age 65 and Over | - 19.39 | - 6.41 | 0.00 | 7.05 | 19.12 | 67 |
| Single Persons | - 13.63 | - 4.81 | - 0.02 | 4.60 | 13.08 | 62 |
| Married Persons | - 8.50 | - 3.11 | - 0.06 | 3.07 | 8.71 | 57 |
| Widowed Persons | - 18.22 | - 7.61 | 0.41 | 8.82 | 19.97 | 58 |
| Divorced Persons | - 21.57 | - 8.48 | 0.61 | 10.55 | 20.73 | 54 |
| Separated Persons | - 22.91 | - 10.45 | 0.30 | 9.52 | 20.13 | 56 |
| **Family Characteristics** | | | | | | |
| Total # of Census Families | - 7.15 | - 2.59 | - 0.02 | 2.78 | 7.48 | 56 |
| Husband-Wife Census Families | - 8.19 | - 2.85 | - 0.02 | 2.95 | 8.18 | 59 |
| Lone Parent Census Families | - 10.41 | - 1.93 | - 0.04 | 0.84 | 9.53 | 82 |
| Census Family Children | - 14.51 | - 5.02 | 0.05 | 5.05 | 14.43 | 63 |
| People in Census Families | - 9.44 | - 3.19 | 0.00 | 3.24 | 9.53 | 61 |
| People Not in Census Families | - 19.36 | - 6.80 | - 0.03 | 5.89 | 18.96 | 68 |
| **Household and Dwelling Characteristics** | | | | | | |
| Owned Dwellings | - 6.75 | - 2.51 | 0.00 | 2.41 | 6.80 | 54 |
| Rented Dwellings | - 13.97 | - 3.79 | 0.00 | 3.68 | 13.58 | 70 |
| Single Detached Dwellings | - 5.94 | - 2.36 | - 0.01 | 2.23 | 5.98 | 40 |
| Apts With Less Than 5 Storeys | - 8.04 | - 3.38 | - 0.51 | 2.07 | 7.25 | 64 |
| Apts With 5 or More Storeys | - 6.44 | - 2.37 | 0.12 | 1.93 | 6.88 | 36 |
| Movable Dwellings | - 12.25 | - 5.52 | 0.80 | 6.84 | 15.26 | 48 |
| All Other Types of Dwellings | - 13.81 | - 5.24 | 0.09 | 6.08 | 14.79 | 55 |
| One Person Households | - 15.74 | - 4.22 | 0.00 | 4.73 | 15.46 | 72 |
| Two Person Households | - 15.78 | - 5.99 | 0.11 | 6.57 | 16.36 | 57 |
| Three Person Households | - 20.29 | - 8.25 | 0.02 | 8.85 | 20.53 | 51 |
| Four or Five Person Households | - 15.91 | - 6.47 | 0.16 | 7.29 | 15.98 | 51 |
| Six or More Person Households | - 25.23 | - 12.60 | - 2.12 | 8.48 | 22.15 | 54 |
| Non Census Family Households | - 15.97 | - 5.15 | 0.00 | 5.14 | 16.05 | 69 |
| One Census Family Households | - 7.10 | - 2.58 | 0.07 | 2.95 | 7.41 | 56 |
| Multiple Census Family Hhlds | - 26.53 | - 13.34 | - 3.13 | 8.55 | 21.56 | 39 |
| Hhld Maintainers Aged < 25 | - 19.76 | - 9.84 | - 1.42 | 6.33 | 16.59 | 59 |
| Hhld Maintainers Aged 25-34 | - 16.94 | - 6.20 | 0.03 | 5.73 | 15.21 | 61 |
| Hhld Maintainers Aged 35-44 | - 15.77 | - 5.70 | 0.12 | 5.50 | 15.01 | 63 |
| Hhld Maintainers Aged 45-64 | - 15.44 | - 5.68 | 0.04 | 5.78 | 14.76 | 58 |
| Hhld Maintainers Aged > 64 | - 15.74 | - 5.77 | 0.02 | 5.66 | 15.70 | 65 |
| Male Households Maintainers | - 7.63 | - 2.94 | - 0.03 | 2.71 | 7.20 | 48 |
| Female Household Maintainers | - 15.84 | - 5.72 | 0.12 | 6.66 | 18.07 | 59 |

## C. Census Tracts (CTs) and Provincial Census Tracts (PCTs)

Table 5 summarizes the level of consistency for all sampled CTs in Canada with a population count greater than 50, and Table 6 summarizes the level of consistency for all sampled PCTs in Canada with a population count greater than 50. Both CTs and PCTs have larger populations on average than CSDs. PCTs have slightly larger populations on average than CTs, however CT boundaries were respected better than PCT boundaries when forming WAs. The consistency for CTs was consequently better than for PCTs for most characteristics, while the consistency for PCTs was better than for CSDs for most characteristics. The characteristics for which this was not true were generally those with poor consistency at all geographic levels. The consistency of the raking ratio estimates was better than for estimates using simple weights for a majority of CTs and PCTs, for almost all characteristics.

## D. Enumeration Areas (EAs)

EAs are the components of WAs. All but five which received special treatment (see Chapter VI, Section A) were only a part of one WA, which is the lowest level at which sample estimates are forced to agree with population counts. Also, the initial weights used were the same for all persons and households in the same WA, even if the sampling fraction differed among the EAs in the WA, whereas the simple weights were calculated at the EA level. Consequently, the consistency at the EA level can not be expected to be as good as at higher levels. Table 7 shows that the consistency for the raking ratio estimates was better than the consistency for estimates using simple weights for less than 20% of all sampled EAs in Canada with a population count greater than 50, for all characteristics studied.

## E. Cells of the Weighting Matrices

The RREP only guarantees that the estimated row and column totals of the cross-classification matrix will agree with the corresponding population counts. There is no control on the individual cells of the matrix. The consistency at the cell level for both the household and person matrices for five randomly selected WAs was studied (only cells with some in-sample units were included). Over all ten matrices, the consistency for the raking ratio estimates was better than the consistency for simple estimates using weights equal to the inverse of the WA household sampling fraction for 58% of the cells, and worse for 42%. The cells for which the raking ratio estimates were better than the simple estimates had larger population counts on average than the remaining cells. Also, the discrepancies tended to decrease as the population counts of the cells increased.

There was a definite tendency for the RREP to over-estimate cells. Over all ten matrices, 58% of the cells were over-estimated while the remaining 42% were under-estimated. Twenty-two percent of the cells with a non-zero population count had no units in the sample. Since the row and column estimates must agree with the population counts, this means that the estimates in the other cells must be increased to make up for the under-estimation in these cells. This probably explains most if not all of the over-estimation. However, because the size of the cells with no sample is small, the over-estimation is also small.

For more information on the Consistency Study, see Rathwell (1990).

**Table 5. Percentiles of Sample Estimate and Population Count Discrepancies (as a Percentage of the Population Count) for CTs and Percentage of Improved CTs**

| Characteristic | Percentiles of Discrepancies | | | | | Percentage of CTs for which Raking Ratio Improved Over Simple Estimates |
|---|---|---|---|---|---|---|
| | 10th | 25th | 50th | 75th | 90th | |
| **Person Characteristics** | | | | | | |
| Males | - 2.13 | 0.00 | 0.00 | 0.00 | 1.93 | 83 |
| Females | - 1.95 | 0.00 | 0.00 | 0.00 | 1.93 | 82 |
| Total Person Population | - 1.67 | 0.00 | 0.00 | 0.00 | 1.52 | 81 |
| Age 0-5 | - 7.69 | - 3.31 | 0.01 | 3.52 | 8.43 | 76 |
| Age 6-14 | - 6.95 | - 2.51 | - 0.02 | 2.44 | 6.69 | 77 |
| Age 15-24 | - 4.85 | - 1.15 | 0.00 | 1.00 | 4.66 | 83 |
| Age 25-34 | - 5.28 | - 2.27 | - 0.02 | 2.12 | 5.57 | 73 |
| Age 35-44 | - 5.64 | - 2.14 | - 0.09 | 2.04 | 5.46 | 74 |
| Age 45-54 | - 6.28 | - 2.00 | - 0.12 | 2.01 | 5.84 | 77 |
| Age 55-64 | - 5.64 | - 2.00 | 0.00 | 2.22 | 6.46 | 79 |
| Age 65 and Over | - 5.17 | 0.00 | 0.00 | 0.00 | 5.27 | 85 |
| Single Persons | - 2.85 | - 0.60 | - 0.04 | 0.63 | 2.82 | 81 |
| Married Persons | - 2.32 | - 0.74 | - 0.07 | 0.60 | 2.01 | 77 |
| Widowed Persons | - 16.22 | - 7.93 | 0.31 | 8.20 | 17.01 | 59 |
| Divorced Persons | - 21.36 | - 10.75 | 0.36 | 11.23 | 21.66 | 55 |
| Separated Persons | - 24.42 | - 12.65 | 0.17 | 11.96 | 24.32 | 57 |
| **Family Characteristics** | | | | | | |
| Total # of Census Families | - 2.03 | - 0.30 | 0.00 | 0.23 | 1.70 | 79 |
| Husband-Wife Census Families | - 2.22 | - 0.33 | 0.01 | 0.25 | 2.05 | 81 |
| Lone Parent Census Families | - 5.68 | - 0.26 | - 0.01 | 0.17 | 6.13 | 86 |
| Census Family Children | - 3.28 | - 0.34 | 0.03 | 0.52 | 3.39 | 84 |
| People In Census Families | - 2.07 | - 0.03 | 0.00 | 0.06 | 1.88 | 84 |
| People Not In Census Families | - 4.37 | - 0.29 | - 0.02 | 0.13 | 4.26 | 84 |
| **Household and Dwelling Characteristics** | | | | | | |
| Owned Dwellings | - 1.81 | - 0.01 | 0.00 | 0.01 | 2.02 | 82 |
| Rented Dwellings | - 2.87 | - 0.01 | 0.00 | 0.01 | 2.73 | 81 |
| Single Detached Dwellings | - 2.90 | - 1.14 | 0.04 | 1.08 | 2.96 | 58 |
| Apts With Less Than 5 Storeys | - 7.43 | - 3.19 | - 0.42 | 1.95 | 6.14 | 52 |
| Apts With 5 or More Storeys | - 7.16 | - 2.74 | 0.01 | 2.85 | 8.24 | 39 |
| Movable Dwellings | - 9.78 | - 4.81 | 0.41 | 5.20 | 12.27 | 46 |
| All Other Types of Dwellings | - 9.80 | - 4.19 | 0.40 | 4.76 | 11.29 | 52 |
| One Person Households | - 4.59 | 0.00 | 0.00 | 0.00 | 4.35 | 85 |
| Two Person Households | - 6.91 | - 2.82 | 0.30 | 3.55 | 7.69 | 66 |
| Three Person Households | - 14.29 | - 7.20 | - 0.41 | 6.58 | 13.31 | 54 |
| Four or Five Person Households | - 11.01 | - 5.34 | 0.06 | 5.07 | 10.44 | 56 |
| Six or More Person Households | - 31.23 | - 18.37 | - 4.06 | 11.43 | 26.77 | 51 |
| Non Census Family Households | - 4.34 | 0.00 | 0.00 | 0.00 | 3.92 | 84 |
| One Census Family Households | - 2.10 | - 0.67 | 0.08 | 0.73 | 2.08 | 73 |
| Multiple Census Family Hhlds | - 35.67 | - 21.14 | - 4.87 | 10.28 | 26.72 | 46 |
| Hhld Maintainers Aged < 25 | - 22.02 | - 11.03 | - 1.15 | 9.01 | 20.36 | 57 |
| Hhld Maintainers Aged 25-34 | - 7.67 | - 3.45 | 0.08 | 3.59 | 7.83 | 68 |
| Hhld Maintainers Aged 35-44 | - 8.45 | - 3.82 | 0.02 | 3.68 | 8.15 | 67 |
| Hhld Maintainers Aged 45-64 | - 6.34 | - 3.01 | 0.07 | 3.03 | 6.34 | 67 |
| Hhld Maintainers Aged > 64 | - 7.95 | - 2.87 | - 0.12 | 2.85 | 7.70 | 74 |
| Male Households Maintainers | - 2.89 | - 1.06 | 0.00 | 1.00 | 2.53 | 68 |
| Female Household Maintainers | - 6.77 | - 2.48 | 0.00 | 2.59 | 7.09 | 70 |

**Table 6. Percentiles of Sample Estimate and Population Count Discrepancies (as a Percentage of the Population Count) for PCTs and Percentage of Improved PCTs**

| Characteristic | Percentiles of Discrepancies | | | | | Percentage of PCTs for which Raking Ratio Improved Over Simple Estimates |
|---|---|---|---|---|---|---|
| | 10th | 25th | 50th | 75th | 90th | |
| **Person Characteristics** | | | | | | |
| Males | - 2.56 | - 1.03 | 0.00 | 1.03 | 2.67 | 70 |
| Females | - 2.71 | - 0.98 | 0.00 | 1.05 | 2.60 | 71 |
| Total Person Population | - 2.11 | - 0.81 | 0.00 | 0.76 | 2.06 | 70 |
| Age 0-5 | - 9.10 | - 3.88 | 0.02 | 4.24 | 8.75 | 70 |
| Age 6-14 | - 7.30 | - 3.34 | - 0.11 | 3.33 | 7.34 | 71 |
| Age 15-24 | - 6.27 | - 2.68 | 0.00 | 2.52 | 6.69 | 73 |
| Age 25-34 | - 6.18 | - 2.85 | - 0.06 | 2.76 | 6.29 | 67 |
| Age 35-44 | - 6.96 | - 2.92 | 0.01 | 3.08 | 6.51 | 70 |
| Age 45-54 | - 8.10 | - 3.52 | 0.05 | 3.54 | 8.00 | 70 |
| Age 55-64 | - 7.73 | - 3.34 | 0.06 | 3.47 | 8.04 | 69 |
| Age 65 and Over | - 7.67 | - 2.79 | 0.00 | 2.75 | 7.08 | 75 |
| Single Persons | - 3.86 | - 1.42 | 0.02 | 1.58 | 3.91 | 69 |
| Married Persons | - 2.51 | - 1.04 | - 0.09 | 0.95 | 2.41 | 70 |
| Widowed Persons | - 14.31 | - 7.61 | - 0.13 | 7.21 | 14.74 | 59 |
| Divorced Persons | - 22.77 | - 11.34 | 1.89 | 13.72 | 25.20 | 53 |
| Separated Persons | - 25.88 | - 14.01 | - 1.05 | 13.12 | 26.44 | 54 |
| **Family Characteristics** | | | | | | |
| Total # of Census Families | - 2.06 | - 0.84 | 0.00 | 0.84 | 2.18 | 70 |
| Husband-Wife Census Families | - 2.40 | - 0.92 | 0.00 | 0.97 | 2.43 | 72 |
| Lone Parent Census Families | - 11.62 | - 4.51 | - 0.05 | 4.20 | 10.81 | 74 |
| Census Family Children | - 4.25 | - 1.55 | 0.07 | 1.77 | 4.23 | 72 |
| People In Census Families | - 2.48 | - 0.98 | 0.00 | 0.95 | 2.58 | 74 |
| People Not In Census Families | - 7.23 | - 2.61 | 0.00 | 2.63 | 7.10 | 77 |
| **Household and Dwelling Characteristics** | | | | | | |
| Owned Dwellings | - 2.18 | - 0.87 | - 0.00 | 0.76 | 2.07 | 67 |
| Rented Dwellings | - 6.58 | - 2.32 | 0.00 | 2.34 | 6.37 | 72 |
| Single Detached Dwellings | - 2.35 | - 1.17 | 0.02 | 1.16 | 2.44 | 51 |
| Apts With Less Than 5 Storeys | - 8.73 | - 3.96 | - 0.46 | 2.77 | 7.15 | 62 |
| Apts with 5 or More Storeys | - 11.49 | - 5.05 | 0.79 | 4.40 | 9.71 | 40 |
| Movable Dwellings | - 17.70 | - 8.38 | 0.59 | 7.61 | 16.63 | 48 |
| All Other Types of Dwellings | - 14.48 | - 6.59 | 0.35 | 6.75 | 15.89 | 55 |
| One Person Households | - 8.38 | - 2.97 | 0.00 | 2.90 | 8.04 | 75 |
| Two Person Households | - 7.94 | - 3.73 | 0.17 | 3.72 | 8.00 | 62 |
| Three Person Households | - 14.18 | - 7.20 | - 0.00 | 6.97 | 14.41 | 54 |
| Four or Five Person Households | - 8.82 | - 4.52 | 0.21 | 4.72 | 9.09 | 56 |
| Six or More Person Households | - 27.93 | - 14.93 | - 1.40 | 12.04 | 27.55 | 51 |
| Non Census Family Households | - 8.14 | - 2.63 | 0.00 | 2.92 | 7.05 | 75 |
| One Census Family Households | - 2.20 | - 0.93 | 0.06 | 1.01 | 2.29 | 68 |
| Multiple Census Family Hhlds | - 40.55 | - 30.78 | - 9.42 | 14.13 | 29.67 | 59 |
| Hhld Maintainers Aged < 25 | - 22.45 | - 11.77 | - 1.14 | 10.04 | 20.82 | 56 |
| Hhld Maintainers Aged 25-34 | - 8.15 | - 3.71 | 0.12 | 4.15 | 8.50 | 64 |
| Hhld Maintainers Aged 35-44 | - 8.76 | - 4.22 | 0.11 | 4.49 | 9.25 | 63 |
| Hhld Maintainers Aged 45-64 | - 6.77 | - 3.23 | 0.19 | 3.41 | 6.76 | 63 |
| Hhld Maintainers Aged > 64 | - 8.77 | - 3.60 | - 0.13 | 3.30 | 7.93 | 69 |
| Male Households Maintainers | - 2.77 | - 1.39 | - 0.04 | 1.16 | 2.54 | 59 |
| Female Household Maintainers | - 8.33 | - 3.49 | 0.11 | 3.98 | 9.07 | 64 |

**Table 7. Percentage of All Sampled EAs in Canada with Population Counts Over 50 for which Raking Ratio Improved Over Simple Weights**

| Characteristic | Percentage of Improved EAs | Characteristic | Percentage of Improved EAs |
|---|---|---|---|
| Females | 7 | Owned Dwellings | 5 |
| Males | 7 | Rented Dwellings | 4 |
| Age 0-5 | 19 | Apts With Less Than 5 Storeys | 2 |
| Age 6-14 | 17 | Apts With 5 or More Storeys | 1 |
| Age 15-24 | 17 | Movable Dwellings | 0 |
| Age 25-34 | 15 | All Other Types of Dwellings | 3 |
| Age 35-44 | 16 | One Person Households | 9 |
| Age 45-54 | 19 | Two Person Households | 16 |
| Age 55-64 | 19 | Three Person Households | 18 |
| Age 65 and Over | 15 | Four or Five Person Households | 12 |
| Single Persons | 10 | Six or More Person Households | 3 |
| Married Persons | 7 | Non Census Family Households | 9 |
| Widowed Persons | 11 | One Census Family Households | 6 |
| Divorced Persons | 10 | Multiple Census Family Hhlds | * |
| Separated Persons | 6 | Hhld Maintainers Aged < 25 | 5 |
| Total # of Census Families | 6 | Hhld Maintainers Aged 25-34 | 16 |
| Husband-Wife Census Families | 7 | Hhld Maintainers Aged 35-44 | 16 |
| Lone Parent Census Families | 8 | Hhld Maintainers Aged 45-64 | 16 |
| Census Family Children | 11 | Hhld Maintainers Aged > 64 | 11 |
| People in Census Families | 7 | Male Households Maintainers | 7 |
| People Not in Census Families | 14 | Female Household Maintainers | 12 |

* There were no EAs with more than 50 multiple census family households.

## F. Consistency of the Mother Tongue Characteristic

A separate study was done on the consistency for responses to the mother tongue question. At the Canada level, the discrepancy between the sample estimate and the population count for both "English only" and "French only" categories was very small. However, rather large discrepancies were noted for multiple responses (- 9.67%), especially for English and French (- 7.35%) and English and non-official languages (- 11.64%).

A higher percentage of people gave multiple responses on the Form 2A (3.9% before imputation) than on the Form 2B (3.3% before imputation), so that the sampling fraction for this category was low. Furthermore, multiple responses were included in the other mother tongue columns of the person weighting matrix, which were frequently collapsed with the English only and French only columns. These two factors, added to the fact that multiple responses made up a relatively small percentage of the population, resulted in a large under-estimation of the number of persons with more than one mother tongue (see Chapter VI, Section B).

The higher percentage of multiple responses on the Form 2A than on the Form 2B was not due to Edit and Imputation or data processing, nor can it be explained by sampling variability. It seems that respondents interpreted the mother tongue question differently on the Form 2B than on the 2A. Although the precise reason for this phenomenon is unknown, one possibility is that additional language and ethnic origin questions on the Form 2B may have helped reduce the number of people who reported more than one mother tongue by providing them with another opportunity to report their other spoken languages and/or origin. Also, the fact that the question instructions, which specifically mentioned that multiple responses were permissible, were a part of the Form 2A and not in a separate booklet as they were for the Form 2B, may have contributed to the different frequencies of multiple responses.

For more information, see Daoust (1988).

## VIII. SAMPLING VARIANCE

Sampling error can be divided into two components: variance and bias. The variance measures the variability of the estimate about its average value in hypothetical repetitions of the survey process, while the bias is defined as the difference between the average value of the estimate in hypothetical repetitions and the true value being estimated. Chapter V presented results of the Sampling Bias Study, describing the nature and extent of bias in the census sample prior to weighting. Chapters VI and VII presented results on the sampling bias following the application of the weighting procedure. Even with a perfectly unbiased sampling method, the results would still be subject to variance, simply because the estimates are based only on a sample. The variance may be estimated using the data collected by the sample survey[12]. The Sampling Variance Study was carried out to estimate the effect of the sampling and estimation procedures on those census figures that are based on sample data.

On the basis of the 2B sample data, thousands of tables are produced by Statistics Canada. Conceptually, a measurement of precision, the estimated sampling variance, can be associated with every estimate calculated in these tables. This measurement takes into account both the sample design and the estimation method. In practice, however, it cannot be calculated for every census estimate because of high data processing costs. Sampling variance is thus estimated for only a subset of census estimates. From this, the combined effect of the sample design and the estimation method on the sampling variance can be estimated. Simple estimates of sampling variance, which are inexpensive to calculate, can then be adjusted for this impact to produce estimates of sampling variance for any census estimates.

The square roots of the sampling variances, known as standard errors, can be approximated using the data in Tables 8 and 9. Table 8 gives non-adjusted (simple) standard errors of census sample estimates. The figures in this table were determined by assuming that 1 in 5 simple random sampling and simple weighting by 5 was used. The standard errors are expressed in Table 8 as a function of the size of both the census estimate and the geographic area. For example, for an estimate of 250 persons in a geographic area with a total of 1,000 persons, the non-adjusted standard error is 25.

---

[12] Unfortunately, the sampling variance does not provide any indication of the extent of non-sampling error.

**Table 8. Non - Adjusted Estimates of Standard Errors of Sample Estimates**

| Estimated Total | Total Number of Persons, Households, Dwellings or Families in the Area | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 500 | 1,000 | 2,500 | 5,000 | 10,000 | 25,000 | 50,000 | 100,000 | 250,000 |
| 50 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 |
| 100 | 18 | 19 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| 250 | 22 | 25 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| 500 | 0 | 30 | 40 | 40 | 45 | 45 | 45 | 45 | 45 |
| 1,000 | | 0 | 50 | 55 | 60 | 60 | 65 | 65 | 65 |
| 2,500 | | | 0 | 70 | 85 | 95 | 95 | 100 | 100 |
| 5,000 | | | | 0 | 100 | 130 | 130 | 140 | 140 |
| 10,000 | | | | | 0 | 150 | 180 | 190 | 200 |
| 25,000 | | | | | | 0 | 220 | 270 | 300 |
| 50,000 | | | | | | | 0 | 320 | 400 |
| 100,000 | | | | | | | | 0 | 490 |
| 250,000 | | | | | | | | | 0 |

| Estimated Total | Total Number of Persons, Households, Dwellings or Families in the Area | | | | | |
|---|---|---|---|---|---|---|
| | 500,000 | 1,000,000 | 2,500,000 | 5,000,000 | 10,000,000 | 25,000,000 |
| 50 | 15 | 15 | 15 | 15 | 15 | 15 |
| 100 | 20 | 20 | 20 | 20 | 20 | 20 |
| 250 | 30 | 30 | 30 | 30 | 30 | 30 |
| 500 | 45 | 45 | 45 | 45 | 45 | 45 |
| 1,000 | 65 | 65 | 65 | 65 | 65 | 65 |
| 2,500 | 100 | 100 | 100 | 100 | 100 | 100 |
| 5,000 | 140 | 140 | 140 | 140 | 140 | 140 |
| 10,000 | 200 | 200 | 200 | 200 | 200 | 200 |
| 25,000 | 310 | 310 | 310 | 320 | 320 | 320 |
| 50,000 | 420 | 440 | 440 | 440 | 450 | 450 |
| 100,000 | 570 | 600 | 620 | 630 | 630 | 630 |
| 250,000 | 710 | 870 | 950 | 970 | 990 | 990 |
| 500,000 | 0 | 1,000 | 1,260 | 1,340 | 1,380 | 1,400 |
| 1,000,000 | | 0 | 1,550 | 1,790 | 1,900 | 1,960 |
| 2,500,000 | | | 0 | 2,240 | 2,740 | 3,000 |
| 5,000,000 | | | | 0 | 3,160 | 4,000 |
| 10,000,000 | | | | | 0 | 4,900 |

Standard errors are given in Table 8 for only a limited number of values for the estimated total and the total number of persons, households, dwellings or families in the area. The following formula may be used to calculate the non-adjusted standard errors for any estimated total for an area of any size:

$$NASE=\sqrt{\frac{4E(N-E)}{N}}$$

where NASE is the non-adjusted standard error, E is the estimated total and N is the total number of persons, households, dwellings or families in the area. For example, for an estimated total of 750 persons in an area with a total of 9,000 persons, the non-adjusted standard error would be:

$$\sqrt{\frac{4(750)(9,000-750)}{9,000}}=52$$

Table 9 provides adjustment factors[13] by which the non-adjusted standard errors should be multiplied to adjust for the combined effect of the sample design and the estimation procedure. To calculate these adjustment factors, a sample of 401 WAs (out of a total of 5,341 WAs) was selected. The sample was allocated among the ten provinces[14] in such a way as to obtain good estimates of the sampling variance at the provincial level without greatly sacrificing the quality of the estimates at the national level. For each WA in the sample, estimates of the sampling variances for raking ratio estimates were calculated for different categories of all of the characteristics[15] given in Table 9. The estimates of sampling variance at the provincial and national levels were obtained by weighting up the WA level estimates. The adjustment factors for each category of each characteristic were calculated by dividing the square roots of these estimates by the non-adjusted standard errors. Adjustment factors were calculated at the provincial and national levels for each characteristic by averaging the adjustment factors for all of its categories. For further information on how these adjustment factors were calculated, see Béland (1990).

To estimate the standard error for a given census sample estimate, the user should determine from Table 9 the adjustment factor applying to the characteristic and multiply this factor by the non-adjusted standard error selected in Table 8. If the characteristic is not identified in Table 9, the user should pick the adjustment factor shown for the "all other" category. For each characteristic in Table 9, adjustment factors are given at the national and provincial level, as well

---

[13] The squares of the adjustment factors are commonly known as "design effects".

[14] The Yukon and Northwest Territories were grouped with British Columbia.

[15] For example, $15,000 - $25,000 was one of the categories for which estimates of sampling variance were calculated for the characteristic "household income".

**Table 9. Standard Error Adjustment Factors at National or Provincial Level and Percentiles of Weighting Area Level Factors**

| Characteristics | National or Provincial Factor | Percentiles of WA Level Factors | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 50 | 75 | 90 | 95 | 99 | 100 |
| **Population Characteristics** | | | | | | | | |
| Age | | | | | | | | |
| . Age groups 0-4, 5-9, 10-14, 15-19, 20-24, 25-29 | 0.18 | 0.05 | 0.19 | 0.29 | 0.35 | 0.49 | 0.52 | 0.60 |
| . Age groups 30-34, 35-44, 45-54, 55-59, 60-64, 5+, 15+ | 0.36 | 0.13 | 0.33 | 0.46 | 0.51 | 0.56 | 0.61 | 0.74 |
| . Age group 65+ | 0.00 | - | - | - | - | - | - | - |
| Sex | 0.00 | - | - | - | - | - | - | - |
| Marital Status | | | | | | | | |
| . Single, married (excluding separated) | 0.25 | 0.04 | 0.23 | 0.31 | 0.42 | 0.49 | 0.55 | 0.68 |
| . Separated, divorced, widowed | 0.88 | 0.55 | 0.84 | 0.98 | 1.06 | 1.15 | 1.20 | 1.33 |
| Highest level of schooling/ Highest degree, certificate or diploma/Total years of schooling | 0.90 | 0.75 | 0.95 | 1.06 | 1.14 | 1.19 | 1.25 | 1.38 |
| Major field of study | 1.20 | 0.84 | 1.16 | 1.22 | 1.28 | 1.35 | 1.43 | 1.51 |
| Mobility status | | | | | | | | |
| . Non-movers | 1.21 | 0.83 | 1.23 | 1.27 | 1.32 | 1.36 | 1.41 | 1.58 |
| . Movers (migrants, non-migrants) | 1.61 | 0.90 | 1.60 | 1.75 | 1.85 | 1.97 | 2.09 | 2.21 |
| Period of immigration | | | | | | | | |
| . before 1946, 1946-1966 | 0.98 | 0.76 | 1.02 | 1.10 | 1.22 | 1.37 | 1.45 | 1.62 |
| . 1967-1977, 1978-1982, 1983-1986 | 1.51 | 0.80 | 1.45 | 1.55 | 1.78 | 1.90 | 2.11 | 2.20 |
| Age at immigration | 1.10 | 0.71 | 1.15 | 1.29 | 1.38 | 1.44 | 1.54 | 1.67 |
| Place of birth | | | | | | | | |
| . Born in Canada | 1.09 | 0.82 | 1.08 | 1.16 | 1.18 | 1.20 | 1.21 | 1.33 |
| . Born outside Canada | 1.35 | 1.11 | 1.34 | 1.43 | 1.60 | 1.67 | 1.75 | 1.91 |
| Immigrant/Non-immigrant population | 1.12 | 0.81 | 1.10 | 1.24 | 1.38 | 1.46 | 1.52 | 1.69 |
| Citizenship | | | | | | | | |
| . Canada, by birth | 1.13 | 0.88 | 1.14 | 1.17 | 1.20 | 1.27 | 1.32 | 1.58 |
| . Other | 1.59 | 1.04 | 1.40 | 1.65 | 1.88 | 1.95 | 2.12 | 2.30 |

**Table 9. Standard Error Adjustment Factors at National or Provincial Level and Percentiles of Weighting Area Level Factors** - Continued

| Characteristics | National or Provincial Factor | Percentiles of WA Level Factors | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 50 | 75 | 90 | 95 | 99 | 100 |
| Ethnic origin | | | | | | | | |
| . English, French | 1.20 | 0.73 | 1.16 | 1.25 | 1.31 | 1.40 | 1.46 | 1.65 |
| . Other | 1.65 | 1.07 | 1.57 | 1.70 | 1.89 | 1.99 | 2.11 | 2.45 |
| Home language | | | | | | | | |
| . English, French, English and French, English and non-official language | 1.12 | 0.50 | 1.09 | 1.35 | 1.75 | 1.89 | 2.09 | 2.19 |
| . Other language groups | 1.76 | 0.99 | 1.68 | 1.89 | 2.01 | 2.20 | 2.41 | 2.66 |
| Official language | | | | | | | | |
| . English, French, English and French | 1.05 | 0.69 | 1.01 | 1.18 | 1.31 | 1.42 | 1.58 | 1.75 |
| . Other language groups | 1.49 | 0.90 | 1.50 | 1.68 | 1.76 | 1.79 | 1.91 | 2.01 |
| Mother tongue - English | | | | | | | | |
| . Newfoundland, Prince Edward Island, Nova Scotia, British Columbia | 0.92 | 0.24 | 0.96 | 1.45 | 1.62 | 1.90 | 2.23 | 2.45 |
| . Quebec | 1.15 | 0.18 | 1.10 | 1.51 | 1.76 | 1.81 | 1.99 | 2.21 |
| . Other provinces | 0.45 | 0.12 | 0.48 | 0.71 | 0.96 | 1.12 | 1.38 | 1.68 |
| . Canada | 0.53 | - | - | - | - | - | - | - |
| Mother tongue - French | | | | | | | | |
| . Quebec | 0.42 | 0.14 | 0.45 | 0.52 | 0.61 | 0.76 | 0.91 | 1.19 |
| . New Brunswick | 0.75 | 0.19 | 0.79 | 0.98 | 1.24 | 1.60 | 1.84 | 2.04 |
| . Other provinces | 1.04 | 0.09 | 1.12 | 1.49 | 1.71 | 1.89 | 2.06 | 2.40 |
| . Canada | 0.77 | - | - | - | - | - | - | - |
| Mother tongue - Other language groups | 1.70 | 0.73 | 1.63 | 2.11 | 2.44 | 2.51 | 2.60 | 2.70 |
| Industry/Occupation | 0.92 | 0.25 | 0.80 | 1.13 | 1.25 | 1.31 | 1.38 | 1.67 |
| Work activity in 1985 | 0.89 | 0.62 | 0.92 | 1.14 | 1.22 | 1.29 | 1.31 | 1.45 |
| Weeks worked in 1985 | 0.94 | 0.68 | 0.99 | 1.18 | 1.29 | 1.33 | 1.39 | 1.69 |
| Hours worked in reference week | 0.83 | 0.63 | 0.85 | 1.01 | 1.14 | 1.19 | 1.24 | 1.51 |
| Year last worked | | | | | | | | |
| . In 1986, in 1985, before 1985 | 0.89 | 0.60 | 0.94 | 0.99 | 1.05 | 1.11 | 1.20 | 1.33 |
| . Never worked | 1.18 | 0.80 | 1.15 | 1.34 | 1.43 | 1.50 | 1.67 | 1.82 |

**Table 9. Standard Error Adjustment Factors at National or Provincial Level and Percentiles of Weighting Area Level Factors** - Continued

| Characteristics | National or Provincial factor | Percentiles of WA level factors | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 50 | 75 | 90 | 95 | 99 | 100 |
| Class of worker | | | | | | | | |
| . Paid workers | 0.72 | 0.56 | 0.75 | 0.86 | 0.93 | 0.95 | 0.98 | 1.09 |
| . Self-employed unincorporated, unpaid family workers | 0.93 | 0.68 | 0.96 | 1.08 | 1.13 | 1.15 | 1.18 | 1.31 |
| Labour force status participation | | | | | | | | |
| . Employed | 0.75 | 0.59 | 0.76 | 0.83 | 0.86 | 0.91 | 0.93 | 1.04 |
| . Unemployed | 1.06 | 0.76 | 1.04 | 1.14 | 1.20 | 1.27 | 1.38 | 1.53 |
| . Not in labour force | 1.25 | 0.91 | 1.30 | 1.43 | 1.50 | 1.58 | 1.63 | 1.84 |
| Major source of income | | | | | | | | |
| . Wages and salaries | 0.65 | 0.42 | 0.67 | 0.80 | 0.85 | 0.87 | 0.92 | 0.99 |
| . Other | 1.05 | 0.71 | 1.00 | 1.12 | 1.17 | 1.20 | 1.24 | 1.48 |
| Disability | | | | | | | | |
| . Limited at home, school and work | 0.94 | 0.69 | 0.96 | 1.11 | 1.29 | 1.34 | 1.42 | 1.69 |
| . Not limited | 0.61 | 0.41 | 0.58 | 0.69 | 0.74 | 0.78 | 0.81 | 0.84 |
| Census family status | | | | | | | | |
| . Husband, wife, child | 0.20 | 0.05 | 0.20 | 0.24 | 0.26 | 0.28 | 0.31 | 0.34 |
| . Lone parent female | 0.45 | 0.14 | 0.43 | 0.51 | 0.55 | 0.61 | 0.68 | 0.81 |
| . Lone parent male, non-member of a census family | 0.68 | 0.35 | 0.65 | 0.79 | 0.89 | 0.99 | 1.14 | 1.32 |
| Economic family status | | | | | | | | |
| . Husband, wife | 0.14 | 0.06 | 0.16 | 0.21 | 0.28 | 0.34 | 0.36 | 0.42 |
| . Lone parent, child | 0.32 | 0.16 | 0.34 | 0.39 | 0.44 | 0.47 | 0.53 | 0.68 |
| . Other family members | 0.74 | 0.24 | 0.70 | 0.84 | 1.03 | 1.09 | 1.18 | 1.31 |
| Number of persons in census family | 0.04 | 0.00 | 0.00 | 0.05 | 0.07 | 0.09 | 0.11 | 0.13 |
| Number of persons in economic family | 0.18 | 0.08 | 0.19 | 0.24 | 0.33 | 0.41 | 0.45 | 0.71 |
| Age of husband, wife, or reference person of economic family | 1.42 | 0.80 | 1.37 | 1.53 | 1.60 | 1.78 | 1.91 | 2.08 |
| All other population characteristics | 1.00 | - | - | - | - | - | - | - |

**Table 9. Standard Error Adjustment Factors at National or Provincial Level and Percentiles of Weighting Area Level Factors** - Continued

| Characteristics | National or Provincial factor | Percentiles of WA level factors | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 50 | 75 | 90 | 95 | 99 | 100 |
| **Household and Dwelling Characteristics** | | | | | | | | |
| Structural type | | | | | | | | |
| . Single detached | 0.33 | 0.05 | 0.35 | 0.55 | 0.67 | 0.75 | 0.89 | 1.08 |
| . Apartment less than 5 storeys | 0.57 | 0.12 | 0.56 | 0.70 | 0.83 | 0.99 | 1.26 | 1.44 |
| . Other | 0.91 | 0.18 | 0.88 | 0.99 | 1.18 | 1.23 | 1.32 | 1.51 |
| Tenure | 0.00 | - | - | - | - | - | - | - |
| Period of construction | 0.78 | 0.61 | 0.75 | 0.82 | 0.89 | 0.99 | 1.24 | 1.49 |
| Main type of heating equipment/Principal heating fuel | 0.87 | 0.18 | 0.86 | 1.04 | 1.12 | 1.25 | 1.32 | 1.47 |
| Central heating equipment | | | | | | | | |
| . With | 0.42 | 0.09 | 0.38 | 0.54 | 0.60 | 0.70 | 0.89 | 1.19 |
| . Without | 0.78 | 0.23 | 0.79 | 0.91 | 1.03 | 1.12 | 1.20 | 1.39 |
| Household size | | | | | | | | |
| . One person household | 0.00 | - | - | - | - | - | - | - |
| . Other | 0.76 | 0.19 | 0.72 | 1.09 | 1.17 | 1.21 | 1.30 | 1.53 |
| Number of rooms | 0.80 | 0.57 | 0.78 | 0.90 | 0.97 | 1.10 | 1.20 | 1.44 |
| Age of household maintainer | | | | | | | | |
| . 25-34, 55-64, 65-74, 75+ | 0.25 | 0.06 | 0.24 | 0.35 | 0.48 | 0.53 | 0.62 | 0.94 |
| . 0-24, 35-44, 45-54 | 0.92 | 0.38 | 0.90 | 1.05 | 1.14 | 1.21 | 1.30 | 1.49 |
| Sex of household maintainer | | | | | | | | |
| . Male | 0.20 | 0.09 | 0.24 | 0.31 | 0.34 | 0.36 | 0.37 | 0.42 |
| . Female | 0.47 | 0.16 | 0.43 | 0.54 | 0.64 | 0.74 | 0.89 | 1.09 |
| Gross rent/Gross rent as a percentage of household income | 0.75 | 0.48 | 0.79 | 0.91 | 0.94 | 0.96 | 1.01 | 1.21 |
| Owner's major payments/ Owner's major payments as a percentage of household income | 0.84 | 0.62 | 0.87 | 0.95 | 1.01 | 1.04 | 1.11 | 1.29 |
| Household Income | 0.75 | 0.51 | 0.73 | 0.82 | 0.90 | 0.95 | 1.03 | 1.17 |
| Value of dwelling | 0.90 | 0.67 | 0.91 | 1.00 | 1.05 | 1.12 | 1.18 | 1.32 |
| Registered condominium | | | | | | | | |
| . Part | 0.63 | 0.18 | 0.59 | 0.84 | 0.93 | 1.11 | 1.30 | 1.48 |
| . Not part | 0.15 | 0.07 | 0.14 | 0.19 | 0.28 | 0.39 | 0.47 | 0.59 |

**Table 9. Standard Error Adjustment Factors at National or Provincial Level and Percentiles of Weighting Area Level Factors** - Continued

| Characteristics | National or Provincial factor | Percentiles of WA level factors | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 50 | 75 | 90 | 95 | 99 | 100 |
| Household type - One family households | | | | | | | | |
| . Without additional persons | 0.22 | 0.05 | 0.20 | 0.27 | 0.33 | 0.36 | 0.40 | 0.56 |
| . With additional persons | 0.50 | 0.20 | 0.48 | 0.61 | 0.72 | 0.74 | 0.79 | 0.90 |
| Household type - Non family households | 0.00 | - | - | - | - | - | - | - |
| Household type - Other | 1.12 | 0.54 | 1.05 | 1.26 | 1.40 | 1.51 | 1.67 | 1.91 |
| All other household and dwelling characteristics | 1.00 | - | - | - | - | - | - | - |
| **Census Family Characteristics** | | | | | | | | |
| Census family structure | | | | | | | | |
| . Husband and wife | 0.20 | 0.09 | 0.21 | 0.26 | 0.29 | 0.33 | 0.36 | 0.42 |
| . Lone parent male | 0.64 | 0.21 | 0.62 | 0.81 | 0.84 | 0.91 | 1.04 | 1.25 |
| . Lone parent female | 0.46 | 0.19 | 0.45 | 0.57 | 0.65 | 0.69 | 0.74 | 0.91 |
| Census family type | | | | | | | | |
| . Primary family | 0.23 | 0.04 | 0.24 | 0.28 | 0.31 | 0.34 | 0.39 | 0.52 |
| . Secondary family | 0.90 | 0.62 | 0.93 | 1.15 | 1.28 | 1.33 | 1.40 | 1.49 |
| Age groups of children at home | 0.78 | 0.40 | 0.70 | 0.91 | 0.98 | 1.09 | 1.19 | 1.45 |
| Labour force activity of husband, wife, or lone-parent | | | | | | | | |
| . Husband, lone-parent, husband and wife in labour force | 0.40 | 0.23 | 0.43 | 0.50 | 0.55 | 0.59 | 0.71 | 0.93 |
| . Wife in labour force | 0.61 | 0.41 | 0.60 | 0.68 | 0.74 | 0.78 | 0.82 | 1.15 |
| . Other | 0.72 | 0.30 | 0.68 | 0.80 | 0.90 | 0.99 | 1.12 | 1.38 |
| Work activity in 1985 of husband, wife or lone parent | | | | | | | | |
| . Worked in 1985 | 0.48 | 0.11 | 0.45 | 0.50 | 0.54 | 0.57 | 0.59 | 0.63 |
| . Did not work in 1985 | 0.93 | 0.60 | 0.90 | 1.04 | 1.18 | 1.26 | 1.30 | 1.43 |
| All other census family characteristics | 1.00 | - | - | - | - | - | - | - |

**Table 9. Standard Error Adjustment Factors at National or Provincial Level and Percentiles of Weighting Area Level Factors** - Concluded

| Characteristics | National or Provincial factor | Percentiles of WA level factors | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 50 | 75 | 90 | 95 | 99 | 100 |
| **Economic Family Characteristics** | | | | | | | | |
| Economic family structure | | | | | | | | |
| . Husband and wife families | 0.29 | 0.13 | 0.30 | 0.36 | 0.48 | 0.56 | 0.68 | 0.91 |
| . Non husband and wife families | 0.56 | 0.35 | 0.50 | 0.66 | 0.81 | 0.90 | 1.06 | 1.28 |
| Mother tongue of family reference person - English | | | | | | | | |
| . Newfoundland, Prince Edward Island, British Columbia | 0.25 | 0.09 | 0.20 | 0.31 | 0.45 | 0.66 | 0.91 | 1.43 |
| . Quebec | 0.49 | 0.25 | 0.47 | 0.50 | 0.69 | 0.83 | 1.05 | 1.53 |
| . Other provinces | 0.18 | 0.07 | 0.19 | 0.22 | 0.24 | 0.27 | 0.31 | 0.49 |
| . Canada | 0.27 | - | - | - | - | - | - | - |
| Mother tongue of family reference person - French | | | | | | | | |
| . Quebec | 0.12 | 0.04 | 0.13 | 0.17 | 0.21 | 0.29 | 0.36 | 0.51 |
| . Other provinces | 0.88 | 0.30 | 0.90 | 1.07 | 1.21 | 1.28 | 1.35 | 1.69 |
| . Canada | 0.40 | - | - | - | - | - | - | - |
| Mother tongue of family reference person - Other than English or French | | | | | | | | |
| . Newfoundland, Nova Scotia | 0.75 | 0.38 | 0.74 | 0.80 | 0.91 | 0.99 | 1.10 | 1.38 |
| . Other provinces | 0.50 | 0.21 | 0.45 | 0.57 | 0.82 | 0.84 | 0.99 | 1.47 |
| . Canada | 0.56 | - | - | - | - | - | - | - |
| All other economic family characteristics | 1.00 | - | - | - | - | - | - | - |

as at the WA level. Unless the area is smaller than a province, the column headed "National or Provincial Factor" should be selected. Adjustment factors for different provinces are given in Table 9 only for cases where they differ significantly from those at the national level. This only occured for the mother tongue characteristics. If an adjustment factor is needed for a census estimate associated with an area smaller than a province, then the percentiles of WA level factors will provide a more accurate value. The percentiles give the spread of all the adjustment factors calculated in the study at the WA level for the different categories of a characteristic. N% of the adjustment factors at the WA level were below the Nth percentile and 100 - N% were above the Nth percentile. For example, 90% of the adjustment factors at the WA level were below the 90th percentile and 10% were above it. The choice of which percentile to use will depend on how conservative the estimate of the standard error is desired to be. For example, using the 100th percentile would provide a very conservative estimate, while using the 75th percentile would provide a somewhat less conservative estimate.

The following rules should be followed when calculating adjusted standard errors:

(a) When determining the standard error of an estimate relating to families or households, the number of families or households in the area, not the number of persons, should be used for selecting the appropriate column in Table 8.

(b) Unless otherwise specified, family characteristics involving husband, wife, lone-parent or family reference person have the same adjustment factors as population characteristics. For example, the adjustment factor for the characteristic "highest level of schooling of husband, wife, or lone parent of a census family" is the same as the population characteristic "highest level of schooling".

(c) For cross-classifications of two or more characteristics, the largest adjustment factor for the characteristics involved should be used.

(d) All the standard error adjustment factors are for estimates of the number of persons, households, dwellings, or families, as opposed to, for example, dollar values. For example, the household income adjustment factors are for estimates of the number of households whose income falls in a certain dollar range, and not for estimates such as average household income.

The following example illustrates how to calculate the adjusted standard errors. Suppose the estimate of interest is the immigrant population in Ontario. The 1986 estimate for this characteristic was 2,081,200. The 1986 Census count for the population of Ontario was 9,001,170. Since neither number is very close to any of the values given in Table 8, the formula given on page 34 to calculate the non-adjusted standard error should be used. In this case the result would be 2,530. From Table 9, the provincial level adjustment factor for the characteristic "immigrant" is 1.12. Consequently, the adjusted standard error for this estimate is
2,530 x 1.12 = 2,834.

The sample estimate and its standard error may be used to construct an interval within which the unknown population value is expected to be contained with a prescribed confidence. The particular sample selected in this survey is one of a large number of all possible samples of the same size that could have been selected using the same sample design. Estimates derived from

the different samples would differ from each other. If intervals from two standard errors below the estimate to two standard errors above the estimate were constructed using each of the different possible estimates, then approximately 19 out of 20 of such intervals would include the value which would have been obtained in a complete census. Such an interval is called a 95% (19 ÷ 20 = 95%) confidence interval. In order to guarantee 95% confidence, however, these intervals must be calculated using the true standard errors of the sample estimates. The adjusted standard errors calculated from Tables 8 and 9 are only estimates of the true standard errors. For sample estimates at the provincial and national level, however, they should be close enough to the true standard errors to calculate approximate 95% confidence intervals of reasonable precision. Below the provincial level, the adjusted standard errors may not be accurate enough for this purpose.

Using the standard error calculated above, an approximate 95% confidence interval for the number of immigrants in Ontario would thus be 2,081,200 ± 2(2,834) or 2,081,200 ± 5,668.

## IX. CONCLUSION

Sampling is now an accepted and integral part of census-taking. Its use can lead to substantial reductions in costs and respondent burden associated with a census, or alternatively, can allow the scope of a census to be broadened at the same cost. The price paid for these advantages is the introduction of sampling error to census figures that are based on the sample. The effect of sampling is most important for small census figures, whether they are counts for rare categories at the national or provincial level or counts for categories in small geographic areas. It should be noted that response errors and processing errors also contribute to the overall error of census figures and it is the same small census figures that are particularly susceptible to the effects of these non-sampling errors. Therefore, even with a 100% census, many small figures would be of limited reliability. As a general rule of thumb for the 1986 Census, figures of size 50 or less that are based on sample data are of very low reliability, while figures up to size 500 tend to have standard errors in excess of 10% of their size.

The procedures for weighting the sample data up to the population level were carried out successfully, and generally achieved the levels of sample estimate and population count consistency anticipated. The poor consistency at the EA level was somewhat surprising, however, despite the fact that the weighting procedures were not designed to control consistency for EAs. Another notable exception was the poor consistency for multiple responses to the mother tongue question. This was apparently due to respondents interpreting the question differently on the Form 2B than on the 2A. A certain amount of bias was detected in the sample counts of many other characteristics as well. This bias was found to have been introduced partly, but not entirely, during data processing and Edit and Imputation. The remaining bias must have been due to one or more factors such as non-response bias, response bias, the selection of a biased sample by the CRs, etc. For most characteristics, however, the weighting procedures corrected for this bias. Sample estimates which remained biased after weighting were for characteristics with small population counts.

Finally, some changes to the weighting methodology are planned for the 1991 Census. The estimation procedures described in this guide have undergone only minor changes since they were introduced in 1971. Since then, there have been significant advances in the development of alternative weighting procedures. There have also been improvements in the programming languages available to implement the weighting algorithms. Consequently, for the 1991 Census, alternatives to the RREP are being examined which, based on research data, are expected to produce more accurate estimates. In addition, the new weighting procedures are being designed to improve sample estimate and population count consistency at the EA level. These improvements should provide significantly more reliable estimates for census users with no increase in costs or respondent burden.

## APPENDIX

## CROSS-CLASSIFICATION WEIGHTING MATRICES

**Table A1. 1986 Census Household Cross-Classification Matrix Rows**

| Household | Household Maintainer | | Number of Persons in Household | Row No. |
|---|---|---|---|---|
| | Sex | Age | | |
| One or More Family Households | Male | 15-24 | =2 | 1 |
| | | | >2 | 2 |
| | | 25-34 | =2 | 3 |
| | | | >2 | 4 |
| | Female | 15-34 | =2 | 5 |
| | | | >2 | 6 |
| | Male | 35-44 | =2 | 7 |
| | | | >2 | 8 |
| | | 45-54 | =2 | 9 |
| | | | >2 | 10 |
| | | 55-64 | =2 | 11 |
| | | | >2 | 12 |
| | Female | 35-64 | =2 | 13 |
| | | | >2 | 14 |
| | Male | ≥65 | =2 | 15 |
| | | | >2 | 16 |
| | Female | ≥65 | | 17 |
| One Person Non-Family Households | Male | 15-34 | | 18 |
| | Female | 15-34 | | 19 |
| | Male | 35-64 | | 20 |
| | Female | 35-64 | | 21 |
| | Male | ≥65 | | 22 |
| | Female | ≥65 | | 23 |
| 2 Or More Person Non-Family Households | Male | | | 24 |
| | Female | | | 25 |

**Table A2. 1986 Census Household Cross-Classification Matrix Columns**

| Dwelling Tenure | Dwelling Type | Column No. |
|---|---|---|
| Owned | Single Detached | 1 |
| | Other | 2 |
| Rented | Apartment | 3 |
| | Other | 4 |

**Table A3. 1986 Census Person Cross-Classification Matrix Rows**

| Sex | Marital Status | Age | Row No. |
|---|---|---|---|
| Male | Never | 0 - 4 | 1 |
| | | 5 - 9 | 2 |
| | | 10 - 14 | 3 |
| | Married | 15 - 19 | 4 |
| | | 20 - 24 | 5 |
| | | 25 - 44 | 6 |
| | | 45 - 64 | 7 |
| | Ever | 15 - 24 | 8 |
| | | 25 - 34 | 9 |
| | Married | 35 - 44 | 10 |
| | | 45 - 54 | 11 |
| | | 55 - 64 | 12 |
| | | ≥65 | 13 |
| Female | Never | 0 - 4 | 14 |
| | | 5 - 9 | 15 |
| | | 10 - 14 | 16 |
| | Married | 15 - 19 | 17 |
| | | 20 - 24 | 18 |
| | | 25 - 44 | 19 |
| | | 45 - 64 | 20 |
| | Ever | 15 - 24 | 21 |
| | | 25 - 34 | 22 |
| | Married | 35 - 44 | 23 |
| | | 45 - 54 | 24 |
| | | 55 - 64 | 25 |
| | | ≥65 | 26 |

**Table A4. 1986 Census Person Cross-Classification**

**Matrix Columns**

| | Family Status | | Mother Tongue | Column No. |
|---|---|---|---|---|
| Family Members | Husband In A Husband-Wife Family | Without Children | English (E) | 1 |
| | | | French (F) | 2 |
| | | | Other (O) | 3 |
| | | With Children | E | 4 |
| | | | F | 5 |
| | | | O | 6 |
| | Parent In A One Parent Family | | E | 7 |
| | | | F | 8 |
| | | | O | 9 |
| | Wife In A Husband-Wife Family | Some Children < 6 Years | E | 10 |
| | | | F | 11 |
| | | | O | 12 |
| | | No Children < 6 Years | E | 13 |
| | | | F | 14 |
| | | | O | 15 |
| | Children In The Families | 0 - 14 Years | E | 16 |
| | | | F | 17 |
| | | | O | 18 |
| | | ≥ 15 Years | E | 19 |
| | | | F | 20 |
| | | | O | 21 |
| Non-Family Members | Person 1 | | E | 22 |
| | | | F | 23 |
| | | | O | 24 |
| | Other Members Of The Household | | E | 25 |
| | | | F | 26 |
| | | | O | 27 |

## REFERENCES

[1] Béland, Y. Results and Methodology of the 1986 Sampling Variance Study, Statistics Canada Internal Report, August, 1990.

[2] Brackstone, G.J. and Hill, C.J. The Estimation of Total Variance in the 1976 Census, Survey Methodology Journal, Vol. 2, No. 2, Statistics Canada, 1976.

[3] Brackstone, G.J. and Rao, J.N.K. An Investigation of Raking Ratio Estimators, Sankhya, Volume 41, Series C, Pt. 2, 1979, pp. 97-114.

[4] Cochran, W.G. Sampling Techniques (Third Edition). New York, John Wiley & Sons, 1977.

[5] Daoust, P. 1986 Weighting Area Formation Evaluation Report, Statistics Canada Internal Report, September 1987.

[6] _____. Analysis of Large 2A/2B Inconsistency for Multiple Responses to Mother Tongue Variable, Statistics Canada Internal Report, March 1988.

[7] Daoust, P. and Bankier, M. Evaluation of Weighting for the 1986 Census of Population, Statistics Canada Internal Report, January 1989.

[8] Deming, W.E. and Stephan, F.F. On a Least Squares Adjustment of a Sample Frequency Table When the Expected Marginal Totals are Known, Ann. Math. Statist., 11, 1940, pp. 427-444.

[9] Dominion Bureau of Statistics. Eighth Census of Canada, 1941, Administrative Report of the Dominion Statistician. Ottawa: King's Printer, 1945.

[10] _____. Ninth Census of Canada, 1951, Vol. XI, Administrative Report. Ottawa: Queen's Printer, 1955.

[11] _____. Sampling in the Census, S.M.S.03.5, March 1968.

[12] _____. 1961 Census of Canada, General Review, Bulletin 7.2-12, Catalogue 99-537. Ottawa: Queen's Printer, 1970.

[13] Fellegi, I.P. Response Variance and its Estimation, Journal of the American Statistical Association, 59 (December 1964), pp. 1016-1041.

[14] Hansen, M.H., Hurwitz, W.N., and Bershad, M.A. Measurement Errors in Censuses and Surveys, Bulletin of the International Statistical Institute, 38 (August 1959), pp. 359-374.

[15] Rathwell, S. Results and Methodology of the 1986 Sampling Bias Study and the 1986 2A/2B Consistency Study, Statistics Canada Internal Report, August 1990.

**REFERENCES** - Concluded

[16] Royce, D. The Use of Sampling in the 1981 Canadian Census, Statistics Canada Internal Report, December 1983.

[17] Statistics Canada. 1971 Census of Canada, General Review, Vol. VI, Part 1, Catalogue 99-740. Ottawa, 1976.

[18] _____. 1976 Census of Canada, Administrative Report, Part 1, Catalogue 99-850. Ottawa, 1980.

[19] _____. 1976 Census of Canada, Quality of Data, Series 1: Sources of Error - Sampling and Weighting, Catalogue 99-844. Ottawa, 1980.

[20] _____. 1981 Census of Canada, Summary Guide: Total Population, Catalogue 99-902. Ottawa, December 1983.

[21] _____. 1986 Census of Canada, Census Handbook, Catalogue 99-104E. Ottawa, June 1988.

[22] _____. User's Guide to the Quality of 1986 Census Data: Coverage, Catalogue 99-135E. Ottawa, March 1990.

[23] Stuart, A. Basic Ideas of Scientific Sampling (Second Edition). New York: Hafner Press, 1976.